Archive of SID

# مروری بر روش‌های مدل‌سازی هم‌پوشانی در الگوریتم‌های انجمن‌یابی شبکه‌های اجتماعی

سید محمّدمهدی صالحی[۱،*]، علی‌اکبر پویان[۲]

**اطلاعات مقاله**

دریافت مقاله: ۱۳۹۶/۱۱/۲۳
پذیرش مقاله: ۱۳۹۷/۰۴/۰۵

**واژگان کلیدی:**
شبکه‌های اجتماعی،
تحلیل شبکه‌های اجتماعی،
انجمن،
انجمن‌یابی،
هم‌پوشانی.

**چکیده**

شبکهٔ اجتماعی شامل تعدادی از اشخاص است که به‌واسطهٔ یک یا چند نوع از روابط یا اشتراکات، با یکدیگر ارتباط دارند. پیدایش و رشد این شبکه‌ها در فضای مجازی و استفادهٔ روزافزون از آن‌ها، باعث شده است تحلیل شبکه‌های اجتماعی به‌عنوان یک حوزهٔ تحقیقاتی پرمخاطب و میان‌رشته‌ای مطرح باشد. هر انجمن شامل چند عضو شبکهٔ اجتماعی است که با توجه به نحوهٔ دسته‌بندی اعضای شبکه تعیین می‌شود. مطالعهٔ انجمن‌ها، مبحثی مهم و کلیدی در تحلیل شبکه‌های اجتماعی است که ویژگی‌ها و الگوهای مشخّص یا مخفی در لایه‌های یک شبکه را نمایان می‌سازد و موجبات درک بهتر ساختار و رفتار آن را فراهم می‌کند. در این مقاله، ابتدا الگوریتم‌های انجمن‌یابی بر اساس رویکرد آن‌ها به شکل‌گیری انجمن و نیز مدل‌سازی مفهوم هم‌پوشانی (امکان تعلق هم‌زمان اعضا به دو یا چند انجمن) در شش دسته شامل روش‌های طیفی و مرکزگرایی، تابع کیفیت (چگالی)، انتشار برچسب، ساختار، نزدیکی و دسته‌بندی یال‌ها، طبقه‌بندی و بررسی شده‌اند. در ادامه، روش‌های انجمن‌یابی روی چهار دادگان با ساختار متفاوت، پیاده‌سازی و مقایسه شده‌اند. نتایج حاصل از معیارهای ارزیابی کیفیت روی چهار دادگان انتخابی، حاکی از آن است که هر روش بر روی توپولوژی و ویژگی خاصی بهتر عمل می‌کند؛ بنابراین هیچ روشی نمی‌تواند به‌عنوان بهترین روش انجمن‌یابی برای تمام شبکه‌ها مطرح باشد.

## ۱- مقدمه

هر شبکهٔ اجتماعی شامل گروهی از انسان‌ها است که از طریق یک یا چند نوع از روابط، به یکدیگر متّصل هستند. از جملهٔ این روابط می‌توان به دوستی، خویشاوندی، تبادلات مالی، همکاری، گرایش‌های هنری، اجتماعی و سیاسی و وجود صفت مشترک (هم‌وطنی، هم‌کیشی و...) اشاره کرد. شکل‌گیری این شبکه‌ها بر اساس نیاز فطری بشر به تعاملات با همنوعان خود بوده است. پیشرفت‌های فناوری در سالیان اخیر و فراگیرشدن استفاده از لپ‌تاپ‌ها و تبلت‌ها و تلفن‌های همراه هوشمند، باعث ایجاد شکل جدید این شبکه‌ها، یعنی «شبکه‌های اجتماعی برخط[۳] » در فضای مجازی شده است [۱].

شبکه‌های اجتماعی برخط، قابلیت‌ها و امکانات بسیاری به کاربران خود عرضه می‌کنند؛ از جمله می‌توان به سهولت استفاده، رایگان بودن، حذف فواصل فیزیکی و جغرافیایی در تشکیل گروه‌ها، اشتراک‌گذاری انواع فایل‌ها و علاقه‌مندی‌ها، جست‌وجو برای یافتن کاربران دارای علایق یا ویژگی‌های مشترک و ساخت پایگاه داده‌ای از کاربران و سوابق فعالیت آنان اشاره کرد. جذابیت‌های متعدد این شبکه‌ها موجب استفادهٔ روزافزون از آن‌ها شده تا آنجا که افراد زیادی، استفاده از این شبکه‌های ارتباطی مجازی را به ارتباطات دنیای حقیقی ترجیح می‌دهند. این امر باعث شده عبارت «شبکه‌های اجتماعی»، عملاً مترادف با «شبکه‌های اجتماعی برخط» به کار برود.

*. پست الکترونیک نویسنده مسئول: mahdi_salehi@shahroodut.ac.ir
۱. دانشجوی دکتری، دانشکده مهندسی کامپیوتر و IT، دانشگاه صنعتی شاهرود
۲. استادیار، دانشکده مهندسی کامپیوتر و IT، دانشگاه صنعتی شاهرود

[3] Online Social Networks (OSN)

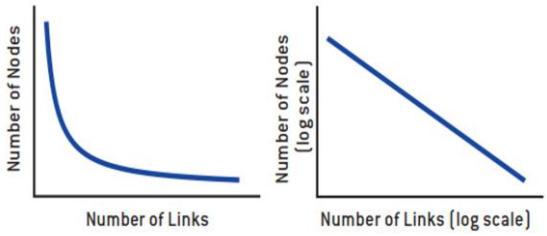

www.SID.ir

Archive of SID

یک شبکهٔ اجتماعی معمولاً با یک گراف $G(V,E)$ مدل می‌شود که در آن، $V$ مجموعه گره‌ها (نودها، رئوس) و $E$ مجموعه یال‌های گراف (اتصالات، لینک‌ها) است. با توجه به حجم داده ناشی از تعداد زیاد گره‌ها (افراد) و یال‌ها (تعاملات میان افراد)، نیازمند استفاده از روش‌های کارآمد در پردازش و مطالعهٔ گراف شبکه‌های اجتماعی هستیم.

«تحلیل شبکه‌های اجتماعی[۱]» دارای کاربردهای بسیاری بوده، به‌عنوان یک مبحث مهم میان‌رشته‌ای در علوم اجتماعی، زیست‌شناسی، اقتصاد، ریاضی، فیزیک و علوم رایانه مطرح است. مطالعهٔ یک شبکه (در دنیای واقعی یا فضای مجازی) برای کاربردهایی چون کشف هسته‌ها، نحوهٔ ایجاد و گردش اطلاعات، ارسال تبلیغات، پیش‌بینی رفتار آتی و ... انجام می‌شود.

تا سال‌ها از شبکه‌های تصادفی[۲] برای مدل کردن مسائل دنیای واقعی استفاده می‌شد. در این شبکه‌ها، احتمال ایجاد یال میان دو گره، با یک تابع توزیع احتمال (عمدتاً نرمال یا پوآسن) بیان می‌شود. این تابع، یال‌ها را بین گره‌ها به‌صورت نسبتاً عادلانه توزیع می‌کند.




شکل ۱: توزیع درجات رئوس شبکه‌های مقیاس آزاد در مختصات خطی و لگاریتمی [۳]

در نقطهٔ مقابل، شبکه‌های مقیاس آزاد[۳] قرار دارند که برای مدل‌سازی رفتار مجموعه‌ها و شبکه‌های پیچیده مطرح شده‌اند [۲]. توزیع یال‌ها در این شبکه‌ها، از نوع «توزیع قانون توانی[۴]» و به فرم $x^{-\alpha}$ است. $\alpha$ بیانگر درجه توانی[۵] (ضریب توانی) توزیع و $x$ بیانگر متغیر توزیع است که مقادیر مختلفی را با توجه به دامنهٔ مسئله (در اینجا: تعداد یال‌ها) اختیار می‌کند. مطابق شکل (۱)، این توزیع در مقیاس خطی، فرم نمایی میرا دارد و در مقیاس لگاریتمی به فرم خط با شیب $-\alpha$ نمایش داده می‌شود. میرایی این توزیع، بیانگر تقسیم ناعادلانهٔ یال‌ها بین گره‌ها در شبکه‌های مقیاس آزاد است؛ بدین معنا که تعداد اندکی از گره‌ها، یال‌های زیادی دارند؛ امّا تعداد زیادی از گره‌ها دارای یال‌های اندکی هستند.

توزیع‌های زیادی برای شبکه‌های مقیاس آزاد وجود دارند؛ مثلاً در توزیع موسوم به zipf[۶]، متغیر $x$ بیانگر فرکانس وقوع یک رویداد در شبکه و رتبه[۷] آن (در مقایسه با دیگر رویدادها) است.

به طور شهودی، یک انجمن[۸] (جامعه، دسته، گروه) در یک شبکهٔ اجتماعی را می‌توان مجموعه‌ای از افراد دانست که علایق مشابهی دارند و ارتباطات بین آن‌ها بیشتر از ارتباطاتشان با افراد خارج از انجمن است. از دید گراف، انجمن‌ها را می‌توان مؤلّفه‌های نسبتاً مستقل گراف یک شبکه دانست که اتصالات میان گره‌های داخلی آن‌ها زیاد بوده، امّا با سایر قسمت‌های شبکه، اتصالات کمتری دارند. انجمن‌ها بازتاب‌دهندهٔ ویژگی‌ها و الگوهای مشخص یا مخفی در لایه‌های یک شبکه هستند. بنابراین بررسی و مطالعهٔ انجمن‌های یک شبکه، باعث درک بهتر آن می‌شود.

مقوله انجمن‌یابی[۹] به‌عنوان یکی از کاربردهای مهم تحلیل شبکه‌های اجتماعی به استخراج اطلاعات یک شبکه از طریق تشخیص انجمن‌های آن و مطالعهٔ این انجمن‌ها می‌پردازد [۴]. از انجمن‌یابی، به کشف اجتماعات یا کشف جوامع نیز یاد می‌شود.

در ساده‌ترین حالت، فرایند انجمن‌یابی شبکه را می‌توان معادل با خوشه‌بندی[۱۰] یا افراز گراف[۱۱] دانست. این حالت، نظیر انجمن‌یابی ناهم‌پوشان[۱۲] است که هر گره را فقط به یک انجمن نسبت می‌دهد. در انجمن‌یابی ناهم‌پوشان، یال‌های بین انجمن‌ها یا «اتصالات ضعیف[۱۳]» که عهده‌دار تبادل اطلاعات میان انجمن‌ها هستند، بررسی نمی‌شوند.

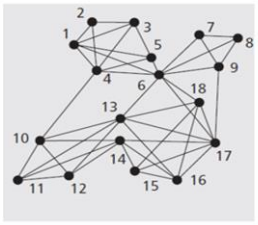
منظور از مفهوم هم‌پوشانی[۱۴] در انجمن‌یابی هم‌پوشان این است که یک گره اجازه تعلق هم‌زمان به دو یا چند انجمن را داشته باشد. شکل (۲) خروجی رویکردهای انجمن‌یابی

[1] Social Network Analysis (SNA)
[2] Random Networks
[3] Scale-free Networks
[4] Power Law Distribution
[5] Degree Exponent
[6] George Kingsley Zipf
[7] Rank
[8] Community
[9] Community Detection (CD)
[10] Clustering
[11] Graph Partitioning
[12] Non-overlapping (Disjoint)
[13] Weak Ties
[14] Overlapping

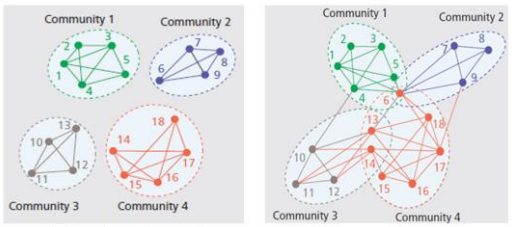

www.SID.ir

Archive of SID

هم‌پوشان و ناهم‌پوشان را برای گراف یک شبکهٔ کوچک نشان می‌دهد. مطابق شکل، رأس ۶ در حالت هم‌پوشان به سه انجمن تخصیص یافته، امّا در حالت ناهم‌پوشان فقط در یک انجمن عضویت دارد. ضمناً نقش اتصالات این رأس با رئوس ۱۳ و ۱۷ و ۱۸ در انجمن‌یابی لحاظ نشده است.



گراف ورودی




انجمن‌یابی ناهم‌پوشان　　انجمن‌یابی هم‌پوشان

شکل ۲: مقایسه انجمن‌یابی هم‌پوشان و ناهم‌پوشان [۵]

مفهوم هم‌پوشانی و پردازش اطلاعات ناشی از آن، یک مزیت در بررسی شبکه‌های اجتماعی محسوب می‌شود؛ زیرا در این شبکه‌ها اشتراکات زیادی بین افراد وجود دارد و امکان عضویت آن‌ها در گروه‌های مختلف برحسب علایق مشترک، بسیار زیاد است.

کاربردهای انجمن‌یابی در علوم مختلف باعث شده است روش‌های متعددی برای این کار ابداع شود. این مقاله به بررسی این روش‌ها و قدرتمندی آن‌ها در انجمن‌یابی روی داده‌گان‌های مختلف می‌پردازد تا مناسب‌ترین روش را با توجه به معیارهای ارزیابی کارایی، مشخص کند.

مهم‌ترین و رایج‌ترین روش‌های انجمن‌یابی با رویکردشان به مفهوم هم‌پوشانی، در بخش (۲) مقاله بررسی می‌شوند. بخش (۳) شامل توضیحات در مورد داده‌گان‌های انتخابی، معرفی معیارهای ارزیابی کارایی و نهایتاً مقایسهٔ عملکرد تعدادی از روش‌های انجمن‌یابی با توجه به این معیارها است. در انتها، بخش (۴) به نتیجه‌گیری و جمع‌بندی مطالب می‌پردازد.

## ۲- مروری بر روش‌های انجمن‌یابی

دسته‌بندی‌های مختلفی روی الگوریتم‌های انجمن‌یابی با رهیافت‌های[1] متفاوت انجام شده است [۶ و ۷]. این مقاله، با تعمیم و به‌روزرسانی ایده‌های مطرح‌شده در [۸]، روش‌های رایج[2] و پرکاربرد انجمن‌یابی را بر اساس رویکرد آن‌ها به نحوهٔ شکل‌گیری انجمن و مدل‌سازی هم‌پوشانی، در شش رده در بخش‌های ۲-۱ تا ۲-۶ طبقه‌بندی کرده است. برخی روش‌های ناهم‌پوشان نیز به دلیل اهمیت یا پیشگام بودن آن‌ها بررسی شده‌اند. نظر به اهمیت مفهوم هم‌پوشانی، بسیاری از روش‌های ناهم‌پوشان برای حالت هم‌پوشان نیز تعمیم یافته‌اند.

ورودی روش‌های انجمن‌یابی، یک گراف ساده (بدون حلقه و یال دوبله) دارای $n$ رأس و $m$ یال است. در گراف‌های خلوت (تنک)[3] پیچیدگی یال‌ها از مرتبه $O(n)$ و تعداد یال‌ها متناسب با تعداد رئوس بوده، امّا در گراف‌های چگال[4] و متراکم، پیچیدگی نزدیک به $O(n^2)$ است.

به‌جز مفهوم هم‌پوشانی، ملاحظات زیر نیز ممکن است برای گراف ورودی در نظر گرفته شود:

- جهت‌دار بودن یال‌های یک گراف
- وزن‌دار بودن یال‌های یک گراف
- نیاز به پارامتر ورودی: عدد پیش‌فرض برای تعداد انجمن‌ها در خروجی یک روش و ...
- پویا بودن شبکه: تکامل آن در طول زمان (ایجاد گره‌ها و یال‌های جدید، حذف گره‌ها و یال‌های قدیمی)
- ورودی چندبعدی: تعریف چند نوع رابطه برای یال‌های گراف
- ورودی چندگانه: گراف دوقسمتی (چندقسمتی) دارای دو (چند) نوع متمایز گره

در جدول ۱ بخش ۷-۲، روش‌های مطرح‌شده برای انجمن‌یابی با ذکر نام، مدل شکل‌گیری انجمن، پیچیدگی زمانی روش، (نا)هم‌پوشانی، امکان کار با گراف وزن‌دار، امکان کار با گراف جهت‌دار، (عدم) نیاز به پارامتر خاص در ورودی روش و سال ارائه روش، فهرست شده‌اند.

### ۲-۱- انجمن‌یابی مبتنی بر طیف و مرکزگرایی

عمده روش‌های انجمن‌یابی که مبتنی بر تحلیل ماتریس‌ها و اطلاعات اولیهٔ توپولوژی یک شبکه‌اند، در این حیطه قرار می‌گیرند.

[1] Approaches
[2] State of the art
[3] Sparse
[4] Dense


www.SID.ir

Archive of SID

روش‌های طیفی[1] ، انجمن‌ها را نواحی پرتراکم یک گراف تعریف می‌کنند و این نواحی را به کمک ماتریس مشابهت گراف و ماتریس‌های حاصل از آن تشخیص می‌دهند. هر مؤلّفه $S_{ij}$ در ماتریس مشابهت[2] $S$ ، وزن شباهت و ارتباط موجود میان رئوس $i$ و $j$ را در این ماتریس متقارن (در گراف غیرجهت‌دار) بیان می‌کند.

ماتریس مجاورت[3] $A$، حالتی خاص از ماتریس مشابهت است که هر مؤلّفه $A_{ij}$ آن، مقدار صفر یا یک دارد. مقدار یک، وجود یال مستقیم بین رئوس $i$ و $j$ را نشان می‌دهد. طیف یک گراف، شامل مقادیر ویژه ماتریس $S$ نظیر آن است. ماتریس $L$ (لاپلاسین نرمالیزه نشده) از رابطه $D-S$ به دست می‌آید که در آن، $D$ ماتریس قطری است و هر مؤلّفه $D_{ii}$ مجموع اوزان متصل به رأس $i$ را بیان می‌کند.

روش‌های طیفی مبتنی بر ماتریس‌های فوق، هم‌پوشان نیستند. تلاش‌هایی برای تعمیم روش‌های طیفی و مقادیر بردارهای ویژه و «تجزیه نامنفی ماتریسی[4] » انجام شده تا هم‌پوشانی نیز محقق شود [۹]. روش Bayesian-NMF مطرح‌شده، به ارائه یک مدل مولد در چارچوب آماری و پارامترهای مشخص با کمک NMF می‌پردازد. روش NMF یک تکنیک استخراج ویژگی و کاهش ابعاد در مباحث یادگیری ماشین است که ماتریس اولیه را به ماتریس‌های نامنفی با ابعاد کوچک‌تر تجزیه می‌کند.

روش موسوم به BIGCLAM[5] از ترکیب NMF و خواص مدل‌های آماری استفاده می‌کند [۱۰]. در این روش برای نمایش ارتباط موجود میان گره‌ها و انجمن‌ها از یک گراف دوقسمتی به نام «گراف وابستگی انجمنی[6] » استفاده می‌شود که دارای دو نوع گره متفاوت است. یک نوع آن، گره‌های گراف اولیه و نوع دیگر آن، خود انجمن‌ها هستند. هر یال این گراف که «یال وابستگی[7] » نامیده می‌شود، بیانگر وجود و میزان تعلق یک گره به یک انجمن است. مفهوم هم‌پوشانی در این روش، با امکان برقراری یال از یک گره به چند انجمن مدل می‌شود. البته این روش نیاز دارد تعداد انجمن‌ها را به‌عنوان پارامتر ورودی دریافت کند.

روش‌های مبتنی بر مرکزگرایی، انجمن‌ها را نواحی متراکم در یک گراف فرض می‌کنند که از طریق یال‌های مرزی با یکدیگر در ارتباط‌اند. یال مرزی، به یالی گفته می‌شود که دو رأس آن به دو انجمن متفاوت تعلق دارد. در مقابل، منظور از یال‌های داخلی یک انجمن، یال‌هایی است که هر دو رأسشان متعلق به همان انجمن باشد. انتظار می‌رود با حذف یال‌های مرزی، آنچه باقی می‌ماند، انجمن‌های گراف به‌عنوان تعدادی زیرگراف مستقل باشد.

معیارهای زیادی (درجه، بینابینی[8] ، رتبه‌صفحه[9] و ...) برای انجمن‌یابی مبتنی بر مرکزگرایی استفاده می‌شوند؛ مثلاً روش ساده GN[10] بر اساس مفهوم «بینابینی یال» عمل می‌کند [۱۱]. برای هر یال گراف، بینابینی یک یال معادل با تعداد کوتاه‌ترین مسیرهای ممکن بین هر دو گره شبکه است که قطعاً از این یال می‌گذرند. این روش ابتدا تمام یال‌های گراف را بر اساس درجه بینابینی آن‌ها مرتب کرده، سپس یال با بزرگ‌ترین درجه بینابینی را حذف می‌کند. مجدداً بینابینی برای گراف باقیمانده محاسبه و یال بعدی حذف می‌شود. این اعمال تا رسیدن به یک شرط توقف (دفعات تکرار، تعداد دلخواه انجمن‌ها) ادامه می‌یابد. روش GN از هم‌پوشانی حمایت نمی‌کند و عملاً زمانی مفید است که انجمن‌ها به‌خوبی تفکیک‌پذیر باشند.

یکی از تعمیم‌های GN، روشی به نام CONGA[11] است که در کنار معیار بینابینی یال از «بینابینی شکافت[12] » نیز استفاده می‌کند [۱۲]. معرفی پارامتر «بینابینی شکافت»، امکان هم‌پوشانی رئوس را فراهم می‌کند. در این روش یک رأس $i$ را به دو رأس فرضی $i_1$ و $i_2$ و یال فرضی بین این دو رأس می‌شکافیم. سپس بینابینی یال را برای این یال فرضی محاسبه می‌کنیم. بهترین حالت را در ارتباط با کلیه حالات ممکن در شکافت یک رأس (با توجه به همسایگانش) در نظر می‌گیریم تا بینابینی شکافت را بیشینه[13] کند. پیچیدگی کار در این است که کدام رأس در چه زمانی و چگونه شکافته شود.

روش DOCNet[14] [۱۳] از دو معیار مرکزگرایی شامل درجه و ضریب خوشه‌بندی[15] (CC) در انجمن‌یابی استفاده می‌کند. ضریب خوشه‌بندی برای گره $i$، به‌صورت

[1] Spectral Clustering
[2] Similarity Matrix
[3] Adjacency Matrix
[4] Nonnegative Matrix Factorization (NMF)
[5] CLuster Affiliation Model for BIG networks (BIGCLAM)
[6] Community Affiliation Graph
[7] Affiliation Edge
[8] Betweenness
[9] PageRank
[10] Girvan-Newman
[11] Cluster Overlap Newman Girvan Algorithm (CONGA)
[12] Split Betweenness
[13] Maximum
[14] Detecting Overlapping Communities in Networks (DOCNet)
[15] Clustering Coefficient (CC)


www.SID.ir

Archive of SID

نسبت یال‌های موجود میان گره‌های همسایه با i، به کل یال‌های ممکن میان آن گره‌ها تعریف می‌شود. این ضریب برای هر گره، بیانگر این است که چقدر گره‌های همسایه با آن گره، خود با یکدیگر همسایه هستند.

به کمک درجه هر گره و ضریب خوشه‌بندی آن، در روش DOCNet پارامتری به نام «اهمیت گره[1] » (NI) تعریف و مقدارش برای هر گره تعیین می‌شود. بر اساس آن، یک هسته کاندیدای اولیه به‌عنوان انجمن مشخص می‌شود. اضافه کردن گره‌های مناسب به هسته، با توجه به NI هر گره انجام خواهد شد.

روش NLA[2] [۱۴] از رتبه‌صفحه استفاده می‌کند. معیار رتبه‌صفحه برای یک گره، بیانگر مقدار احتمال ورود به آن گره، با شروع از دیگر گره‌های گراف در یک پیمایش غیرتصادفی است. در شروع، تخصیص مقدار احتمال اولیه به‌طور مساوی برای هر گره گراف (یک عدد کوچک‌تر از ۱ و نزدیک به آن) انجام می‌شود. با انجام مراحل متوالی و پس از رسیدن به یک شرط توقف، رتبه گره‌ها در گراف بر اساس اهمیتشان، مشخص شده است.

در NLA، به هر گره بر اساس رتبه‌صفحه آن، یک عدد جرم اختصاص می‌یابد و کل گراف به یک میدان پتانسیل سه‌بعدی، نگاشته[3] می‌شود. اختصاص گره‌ها به انجمن بر اساس جرم و مکان قرارگیری گره‌ها در این نگاشت، انجام خواهد شد. این روش در کار با گراف‌های کوچک، سرعت زیادی دارد؛ ولی برای گراف‌های بزرگ مناسب نیست.

## ۲-۲- انجمن‌یابی مبتنی بر تابع کیفیت

برخی از روش‌ها، انجمن را موجودیتی تعریف می‌کنند که مقدار یک تابع کیفیت (تابع چگالی) را بیشینه می‌سازد. این تابع برای هر حالتی از افراز گره‌ها در انجمن‌یابی، مقداری را تولید می‌کند که بیشتر بودن این مقدار، به معنای مطلوب‌تر بودن کیفیت انجمن‌یابی انجام‌شده است.

یکی از متداول‌ترین توابع در انجمن‌یابی، تابع پیمانه‌ای (ماژولاریتی) است [۱۵]. انجمن‌یابی با تابع پیمانه‌ای که به روش CNM[4] یا fast greedy نیز مشهور است، میزان ارزشمندی یک حالت انجمن‌یابی پیشنهادی را به کمک تابع بیان‌شده در رابطه (۱) و در قالب یک الگوریتم سلسله‌مراتبی تجمعی، به‌صورت کمّی برآورد می‌کند:

$$Q = \frac{1}{2m} \sum_{ij} \left[ A_{ij} - \frac{\deg(i).\deg(j)}{2m} \right] \delta(c_i, c_j) \qquad (۱)$$

در رابطه (۱)، m تعداد یال‌ها را بیان می‌کند و $A_{ij}$ بیانگر وجود یال بین دو رأس i و j در ماتریس مجاورت گراف است. $\deg(i)$ بیانگر درجه رأس i و ci بیانگر برچسب انجمنی است که رأس i به آن تعلق دارد. اگر $c_i$ و $c_j$ متعلق به یک انجمن باشند، $\delta(c_i, c_j)$ برابر با یک است و در غیر این صورت، برابر با صفر خواهد بود.

برآورد کیفیت انجمن‌یابی انجام‌شده از مقایسه خروجی گراف در هر حالت خاصی از انجمن‌یابی، با چگالی یک «مدل مرجع[5] » انجام می‌شود. منظور از مدل مرجع، یک گراف تصادفی با همان تعداد رئوس و یال‌های گراف اولیه است. کسر $\deg(i).\deg(j)/2m$ بیانگر احتمال وجود یال میان دو رأس در گراف با توزیع تصادفی یکنواخت است. گراف یک شبکه تصادفی به دلیل توزیع نسبتاً عادلانه یال‌ها ساختار انجمنی ندارد؛ بنابراین چگالی قسمت‌های مختلف آن، اختلاف زیادی با یکدیگر ندارند. با این تعبیر، افزایش تفاوت در مقایسه با ساختار یک گراف تصادفی به معنای رسیدن به حالت مطلوب‌تری در انجمن‌یابی است.

تابع Q برای هر خروجی یک الگوریتم انجمن‌یابی، قابل محاسبه است. این تابع، انجمن‌یابی را به‌صورت ناهم‌پوشان انجام می‌دهد؛ اما به دلیل سهولت استفاده، کاربرد زیادی دارد؛ تا آنجا که حتی از این تابع به‌عنوان معیار ارزیابی کارایی در دیگر روش‌ها نیز استفاده می‌شود.

در محاسبهٔ تابع پیمانه‌ای، روشی به نام Louvain [۱۶] فضای جست‌وجو را کاهش داده، سریع‌تر از روش CNM عمل می‌کند. در ابتدای کار، همهٔ رئوس در انجمن‌های مجزا قرار داده می‌شوند (تعداد انجمن‌ها به اندازهٔ تعداد رئوس). سپس برای هر گره i، مقدار پیمانه‌ای وزن‌دار برای حالتی که گره i به هرکدام از همسایگانش (مثلاً گره j) ملحق شود و یک انجمن جدید بسازد، محاسبه می‌شود. در صورت افزایش در مقدار پیمانه‌ای، عمل الحاق به گره‌ی انجام می‌شود که بیشترین مقدار پیمانه‌ای برای آن به دست آمده است.

تلاش‌هایی برای تعمیم تابع پیمانه‌ای ارائه شده است تا از مفهوم هم‌پوشانی حمایت کند. یک ایده، متغیر احتمالاتی

[1] Node Importance (NI)
[2] Node Location Analysis (NLA)
[3] Mapped
[4] Clauset, Newman and Moore (CNM)
[5] Ground (Null) Model


www.SID.ir

Archive of SID

$a_{i,c}$ را برای بیان میزان تعلق رأس i به انجمن c، مطابق با رابطه (۲) تعریف می‌کند [۱۷]:

$$a_{i,c} = \frac{\sum_{k \in c} A_{ik}}{\sum_{c' \in C_i} \sum_{k \in c'} A_{ik}}$$

$$\forall i \in V, \forall c \in C: \quad 0 \le a_{i,c} \le 1 , \sum_{c \in C} a_{i,c} = 1 \qquad (۲)$$

در اینجا منظور از c، هر انجمن دلخواه و منظور از C یک آرایش از تمام انجمن‌ها (مطابق با انجمن‌یابی پیشنهادی) برای گراف ورودی است. برای هر دو حالت هم‌پوشان و ناهم‌پوشان، اجتماع همهٔ انجمن‌ها برابر با C ( $Uc = C$ ) خواهد بود؛ امّا برای حالت ناهم‌پوشان، اشتراک دو انجمن متمایز حتماً تهی است. برای حالت هم‌پوشان، $C_i$ بیانگر انجمن‌هایی است که گره i عضو آن‌ها است. منظور از 'c'، تمام انجمن‌ها به‌جز انجمن c است ( $c' = C - c$ ).

در ابتدای کار، فرض بر این است که هر گره با میزان احتمال یکسان، به انجمن‌هایی که در آن‌ها عضو است، تعلق داشته باشد. به کمک این متغیر، رابطه (۱) تعمیم داده می‌شود تا ضابطهٔ تابع پیمانه‌ای در حالت هم‌پوشان مطابق با رابطهٔ (۳) تعریف شود:

$$Q_{ov} = \frac{1}{2m} \sum_{c \in C} \sum_{i,j \in c} \left[ A_{ij} - \frac{\deg(i).\deg(j)}{2m} \right] a_{i,c} a_{j,c} \qquad (۳)$$

مقدار $Q_{ov}$ با تعداد انجمن‌هایی که هر گره می‌تواند به آن‌ها متعلق باشد و نیز سهم (قدرت) عضویت در هر انجمن، نسبت مستقیم دارد.

توابع کیفیت دیگری نیز قابل تعریف و استفاده هستند. روش CIS[۱] [۱۸] انجمن را معادل با زیرگرافی می‌داند که به‌صورت محلی، یک تابع چگالی یال را بیشینه می‌کند. مطابق این تعریف، یک انجمن معادل با مجموعه‌ای است که چگالی ارتباطات میان گره‌هایش از یک حدآستانه بیشتر است؛ یعنی اگر یک گره به این مجموعه اضافه یا از آن کم شود، چگالی ارتباطات کاهش می‌یابد. مجموعه زیرگراف‌هایی که این تابع تعریف‌شده را بیشینه می‌کنند ممکن است دارای دو یا چند رأس مشترک باشند، بنابراین می‌توان مفهوم هم‌پوشانی را مدل کرد. قبل از هر بار اجرای الگوریتم، همبندی انجمن‌ها کنترل می‌شود تا اگر یک انجمن به دو یا چند مؤلّفه تجزیه شده بود، مؤلّفه‌ای را حفظ کند که بیشترین چگالی را دارد. تابع کیفیت طبق رابطهٔ (۴) است:

$$f(c) = \frac{w_{in}^c}{w_{in}^c + w_{out}^c} + \lambda_{ep} \qquad (۴)$$

پارامتر $\lambda$ رفتار الگوریتم را در نواحی کم‌تراکم شبکه با دخالت دادن احتمال وجود یال، کنترل می‌کند. پارامتر $w_{in}^c$ مجموع وزن‌های یال‌های داخلی انجمن c را نشان می‌دهد و پارامتر $w_{out}^c$ بیانگر مجموع وزن‌های یال‌های خارجی (مرزی) انجمن c است. برای گراف غیر وزن‌دار، وزن یال‌ها همان تعداد یال‌ها خواهد بود؛ زیرا وزن هر یال برابر با ۱ در نظر گرفته می‌شود. اضافه کردن هر رأس به انجمن، نیازمند برقراری تعادل میان تغییرات چگالی درجات داخلی و تغییرات چگالی یال است.

روش LFM[۲] مکانیزمی مشابه روش‌های فوق دارد [۱۹]. در این روش، بعد از یافتن یک انجمن، گرهی را که هنوز به هیچ انجمنی تخصیص نیافته، به‌طور تصادفی انتخاب کرده، امکان افزودن آن به انجمن فعلی را با توجه به تابع چگالی بررسی می‌کنیم. تابع چگالی این روش مطابق رابطهٔ (۵) تعریف می‌شود:

$$f(c) = \frac{w_{in}^c}{(w_{in}^c + w_{out}^c)^{\gamma}} \qquad (۵)$$

متغیر $\gamma$ در این تابع، یک پارامتر تفکیک است که اندازهٔ انجمن‌های حاصل از این روش را کنترل می‌کند.

رویکرد برخی روش‌ها، تلاش برای بیان تابع کیفیت در قالب مدل‌های آماری (شبکه‌های باور بیزین[۳] ، مدل‌های استنتاجی و ...) است. در این حیطه می‌توان به روش OSLOM[۴] [۲۰] اشاره کرد. این روش، انجمن را ساختاری تعریف می‌کند که در مقایسه با یک مدل مرجع، از لحاظ آماری معنادار[۵] باشد. منظور از مدل مرجع، گراف تصادفی متناظر با گراف شبکه فعلی است و منظور از معناداری آماری، مقدار احتمال یافتن ساختار مشابه با انجمن در حال بررسی (در تعداد رئوس، دنباله درجات، اتصالات داخلی انجمن) نسبت به مدل مرجع. فرایند کار، شامل شروع از یک گره و بررسی امکان بسط آن به‌صورت حریصانه با کنترل میزان معناداری است.

[1] Connected Iterative Scan (CIS)
[2] Lancichinetti and Fortunato Method (LFM)
[3] Bayesian Belief Networks (BBN)
[4] Order Statistics Local Optimization Method (OSLOM)
[5] Significant


www.SID.ir

Archive of SID

روش FOCS[1] جست‌وجو در شبکه‌های بزرگ را بر اساس امتیازات معیارهایی که به‌صورت محلی محاسبه می‌شوند، انجام می‌دهد [۲۱]. این روش دو معیار را معرفی می‌کند که خواص انجمن‌های دنیای واقعی را بازتاب می‌دهند. معیار «همبندی انجمنی[2] » برای هر گره، بیانگر میزان تعلق آن گره به انجمنی است که در آن عضویت دارد و معادل نسبت تعداد همسایگان آن گره که عضو انجمن هستند، به تمام گره‌های انجمن، تعریف می‌شود. گره‌های با بیشترین میزان تعلق، قطعاً در انجمن فعلی می‌مانند. معیار «همبندی همسایگی[3] » برای هر گره، بیانگر میزان دلبستگی آن گره به عضویت در یک انجمن دیگر (با حفظ عضویت در انجمن فعلی) است و معادل نسبت تعداد همسایگان آن گره که عضو انجمن هستند، به تمام همسایگان آن گره، تعریف می‌شود. گره‌های با بیشترین میزان دلبستگی برای پیوستن به یک انجمن دیگر، جزء گره‌های هم‌پوشان خواهند بود.

یک تابع کیفیت را می‌توان برای بررسی عملکرد طیف وسیعی از الگوریتم‌ها به کار برد که بر اساس یک آرایش انتخابی از موجودات (گره‌ها) کار می‌کنند؛ مثلاً انجمن‌یابی شبکه‌های اجتماعی را می‌توان با یک مسئله بهینه‌سازی در الگوریتم‌های فرامکاشفه‌ای[4] (فراکاوشی) [۲۲] مدل کرد. این الگوریتم‌ها دارای رویکرد مکاشفه‌ای[5] یا اتفاقی[6] هستند و برای همه مسائل، جوابی قابل‌قبول (و نه الزاماً بهینه) پیدا می‌کنند.

این الگوریتم‌ها غیرقطعی‌اند و برای هر بار اجرا جواب متفاوتی می‌دهند. معمولاً باید میانگین نتایج تعداد دفعات زیادی از اجرای الگوریتم را در نظر گرفت؛ امّا وقتی فضای جست‌وجوی مسئله خیلی وسیع باشد، انتخاب مناسب برای حل مسئله به شمار می‌روند.

به‌عنوان یک رده مهم از روش‌های فرامکاشفه‌ای، می‌توان به الگوریتم‌های تکاملی اشاره کرد. این الگوریتم‌ها مبتنی بر شبیه‌سازی فرایندهای بیولوژیکی انجام‌شده توسط یک توده یا گروه از موجودات زنده هستند که در طول سال‌ها به تکامل رسیده‌اند.

در روش الگوریتم ژنتیک[7] که پراستفاده‌ترین روش تکاملی محسوب می‌شود، با شروع از یک جمعیت اولیه، مقدار تعیین‌شده توسط یک تابع برازش[8] را برای هرکدام از اعضای جمعیت می‌یابیم. ایجاد تغییرهای پیش‌بینی‌شده روی نمونه‌های مناسب در نسل فعلی برای رسیدن به نسل بعدی، شامل بازتولید با انتخاب والدهای مناسب از نسل فعلی به کمک عملگرهای برش[9] و جهش[10] خواهد بود. نهایتاً تکرار این عملیات به مقدار لازم (رسیدن به حد قابل‌قبولی از برازش، عدم تغییر بعد از چند تکرار و ...) انجام می‌شود.

برای مدل‌کردن مسئله انجمن‌یابی، هر گره شبکه معادل یک عنصر از جمعیت اولیه در نظر گرفته می‌شود و تابع برازش در نقش تابع کیفیت عمل می‌کند. روشی به نام GA-Net[11] [۲۳] و نیز روش تعمیم‌یافتهٔ آن به نام MOGA-Net[12] [۲۴] از ویژگی‌های روش‌های تکاملی برای انجمن‌یابی استفاده می‌کنند.

تلاش‌هایی در تلفیق مفاهیم «تئوری بازی[13]» با روش‌های تئوری گراف انجام شده است. مدل‌کردن انجمن‌یابی هم‌پوشان در شبکه‌های اجتماعی با یک بازی استراتژیک، معادل این است که هر گره را یک عامل حریص بدانیم که ملحق شدن به یک انجمن یا خارج شدن از آن را با میزان سودمندی این عمل می‌سنجد. سودمندی، معادل بیشینه‌سازی مقدار یک تابع بهره (سود)[14] در نظر گرفته می‌شود [۲۵].

در روش پیشنهادی [۲۶] با عنوان Game، متناظر با هر گره، یک تابع بهره و یک تابع اتلاف[15]، تعریف شده است. هر گره (عامل) برای رسیدن به نقطهٔ تعادل میان مقدار بیشینه انتخاب‌شده برای تابع بهره و مقدار کمینه[16] انتخاب‌شده برای تابع اتلاف، می‌تواند در دو یا چند انجمن عضو باشد. در این روش، تابع کیفیت (چگالی) همان تابع بهره یا تابع اتلاف است.

### ۲-۳- انجمن‌یابی مبتنی بر انتشار

انتشار اطلاعات میان افرادی که عضو یک گروه هستند

[1] Fast Overlapped Community Search (FOCS)
[2] Community Connectedness
[3] Neighborhood Connectedness
[4] Meta-heuristic
[5] Heuristic
[6] Stochastic
[7] Genetic Algorithm
[8] Fitness Function
[9] Cross-over
[10] Mutation
[11] Genetic Algorithm for Networks (GA-Net)
[12] Multi-Objective GA-Net (MOGA- et)
[13] Game Theory
[14] Gain (Utility) Function
[15] Loss Function
[16] Minimum


www.SID.ir

Archive of SID

سریع‌تر از انتشار آن میان افرادی که گروه مشترکی ندارند، انجام می‌شود. با استفاده از این ایدهٔ ساده، می‌توان انجمن‌ها را با ردیابی نحوهٔ انتشار اطلاعات در شبکه یافت. بدین معنا که یک انجمن را شامل رئوسی دانست که در انتشار اطلاعات یا خاصیت تعریف‌شده در یک شبکه، تأثیر یکسان می‌گیرند و رفتار مشابهی بروز می‌دهند.

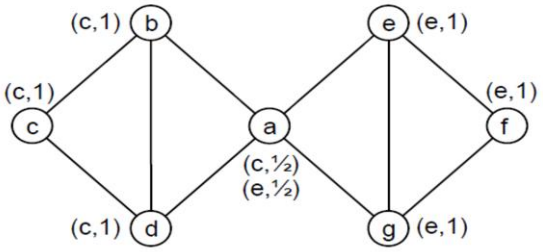
کاربردهای فراوانی برای این گروه از روش‌ها وجود دارد که می‌توان به مواردی چون بررسی اثر «دهان‌به‌دهان[1] » در انتشار اطلاعات و رفتار، نیّت‌کاوی[2] (پردازش اطلاعات مربوط به نظرات افراد)، رفتار و میزان برون‌گرایی اعضای شبکه اجتماعی [۲۷]، شناسایی افراد خبره و «گره‌های تأثیرگذار[3] » در انتشار اطلاعات خاص [۲۸] و انواع بازاریابی [۲۹] خصوصاً «بازاریابی ویروسی[4] » اشاره کرد.

معمولاً نمی‌توان برای گراف متراکمی که انجمن‌های آن به‌سادگی قابل تشخیص نیستند، از روش‌های مرکزگرایی (حذف یال‌های مابین انجمن‌ها) استفاده کرد. در این‌گونه موارد، روش‌های مبتنی بر انتشار، عملکرد بهتری دارند.

یک روش پیشگام و پرکاربرد در این حیطه، «الگوریتم انتشار برچسب[5] » یا LPA نام دارد [۳۰]. ایدهٔ این روش مبتنی بر توزیع اولیهٔ تعدادی برچسب میان گره‌های شبکه است. در ادامه، هر گره برچسبی را اخذ می‌کند که بیشترین تکرار میان همسایگانش را دارد. قسمت‌های چگال شبکه تا حد امکان بسط می‌یابند و به برچسب‌های مشترک می‌رسند. در پایان کار، هر انجمن شامل تعدادی گره با برچسب یکسان (متفاوت با برچسب دیگر انجمن‌ها) خواهد بود. این روش، نیازی به پارامتر ورودی ندارد.

یک چالش مهم در انتشار برچسب، حالتی است که دو یا چند گره از همسایگان دارای برچسب با بیشترین فرکانس تکرار باشند. ساده‌ترین راه برای حل این مشکل، مکانیزم tie-breaking است؛ بدین معنا که برچسب یکی از این همسایگان به‌طور تصادفی انتخاب شود و مراحل الگوریتم ادامه یابد. امکان اخذ برچسب‌های متفاوت در حین اجرای روش باعث می‌شود LPA یک روش غیرقطعی باشد. اجرای چندبارهٔ الگوریتم، راهکاری است که برای این مورد توصیه می‌شود. این روش، سریع بوده، به‌عنوان یک روش مناسب در کار با گراف‌های کم‌تراکم مطرح است. مشخص است که این روش از هم‌پوشانی حمایت نمی‌کند؛ زیرا هر گره فقط یک برچسب دارد؛ بنابراین فقط به یک انجمن تعلق خواهد داشت.

یکی از این تعمیم‌های LPA برای رسیدن به هم‌پوشانی، روش COPRA[6] است [۳۱]. برای تحقق هم‌پوشانی، هر گره با زوج مرتب‌های $(x_1, x_2)$ توصیف می‌شود که در آن، $x_1$ بیانگر شناسهٔ یک انجمن خاص است و $x_2$ ضریب تعلق (بین ۰ و ۱) به آن انجمن را بیان می‌کند. مطابق شکل (۳)، تعلق ۷ گره در یک گراف ساده به دو انجمن c و e دیده می‌شود. گره a به هر دو انجمن تعلق دارد.




شکل ۳: خروجی الگوریتم انتشار برچسب هم‌پوشان [۳۱]

در ابتدای این روش، به هر گره یک برچسب منحصربه‌فرد اختصاص می‌یابد و در هر مرحلهٔ اجرا، ضرایب تعلق هر گره بر اساس میانگین ضرایب همسایگانش تغییر می‌کند. اجرای این روش تا رسیدن به هم‌گرایی و عدم‌تغییر برچسب‌ها ادامه می‌یابد. خروجی این روش نیز به دلیل استفاده از مکانیزم انتخاب تصادفی، غیرقطعی است و نیاز به تکرار در دفعات اجرا دارد.

دیگر روش هم‌پوشان در انتشار برچسب، روش SLPA[7] است که با ایدهٔ نحوهٔ مکالمهٔ افراد با یکدیگر (شنیدن و تصمیم‌گیری بر اساس سابقه ذهنی) ارائه شده است [۳۲]. در این روش، هر گره، حافظه‌ای دارد که سابقهٔ برچسب‌ها را نگهداری می‌کند. این روش همچنین دارای دو پارامتر ورودی است که باید مقداردهی اولیه شوند: حداکثر تعداد مراحل اجرا و حدآستانه مرحله نهایی (پس‌پردازش).

در هر مرحله، یک گره به‌عنوان شنونده انتخاب می‌شود. هر یک از همسایگان گره شنونده، به‌صورت تصادفی یک برچسب با احتمال متناسب با فرکانس وقوع آن در حافظه خودش را برای گره شنونده ارسال می‌کند. گره شنونده

[1] Word of Mouth
[2] Opinion Mining
[3] Influential Nodes
[4] Viral Marketing
[5] Label Propagation Algorithm (LPA)
[6] Community Overlap PRopagation Algorithm (COPRA)
[7] Speaker- Listener Propagation Algorithm (SLPA)


www.SID.ir

Archive of SID

پرتکرارترین برچسب دریافت‌شده را به حافظهٔ خود اضافه می‌کند. نهایتاً در مرحلهٔ پس‌پردازش، انجمن‌ها با اعمال حدآستانه تولید می‌شوند. روش SLPA برای گراف‌های کوچک، دارای سرعت بالایی است.

روشی به نام DEMON[1] هم‌پوشانی را در انجمن‌یابی شبکه‌های مقیاس‌آزاد، مدنظر قرار می‌دهد [۳۳]. ایدهٔ کلی بدین شرح است: هر گره دید محدودی از کل شبکه دارد که شامل انجمن‌های عضو در آن‌ها و نیز انجمن‌های مجاور (محلی) است. با ادغام این دیدگاه‌ها در یک رویکرد دموکراتیک می‌توان به ساختارهای هم‌پوشان رسید؛ یعنی عملاً خود گره‌ها در مورد ساختارهای انجمنی موجود قضاوت می‌کنند. برای تعبیر دیدگاه هر رأس دلخواه$i$ نسبت به کل گراف، یک شبکه Ego برای آن رأس تعریف می‌شود که معادل با زیرگرافی شامل رأس $i$ و رئوس مجاور آن و یال‌های موجود بین این رئوس است.

گروهی از روش‌ها، انتشار برچسب را با یک مسئلهٔ استنتاج آماری مدل می‌کنند. تابع تعلق این روش‌ها به جای یک عدد ساده، مشتمل بر برداری از احتمالات است. $P(X,Y)$ احتمال وجود یال میان گرهی از انجمن $X$ با گرهی از انجمن $Y$ را بیان می‌کند و با میانگین وزنی $P(X,X)$ و $P(Y,Y)$ به دست می‌آید. مکانیزم معروف به MMSB[2] از فرآیندهای اتفاقی استفاده می‌کند [۳۴]. در شروع کار، هر گره با یک احتمال مشخص به هر انجمن تعلق دارد و در ادامه، بردار احتمال متعلق به هر گره از بردارهای سایر گره‌ها تأثیر می‌پذیرد.

در روش MOSES[3] هر گره با یک بردار پنهان[4] شامل تعدادی متغیر بولی مستقل، مدل می‌شود [۳۵]. بسط انجمن‌ها در این روش با اضافه کردن یال‌ها از طریق یک تابع بهینه‌ساز محلی به‌صورت حریصانه انجام می‌شود که تعداد یال‌های قابل بسط در هر مرحله را تعیین می‌کند.

در یک روش جدید به نام NISE[5] از ایدهٔ مؤثر تورم[6] همسایه‌ها استفاده می‌شود [۳۶]. در این روش انتشار، تأکید فقط روی یک گره انتخاب‌شده به‌عنوان نقطهٔ شروع (بذر) نیست، بلکه همسایگان گره نیز در فرایند انتشار دخیل‌اند و یک ناحیهٔ بذر را تشکیل می‌دهند.

[1] Democratic Estimate of the Modular Organization of a Network (DEMON)
[2] Mixed Membership Stochastic Blockmodels (MMSB)
[3] Model-based Overlapping Seed ExpanSion (MOSES)
[4] Latent Vector
[5] Neighborhood-Inflated Seed Expansion (NISE)

### ۲-۴- انجمن‌یابی مبتنی بر ساختار

گروهی از روش‌ها، انجمن را یک ساختار دقیق از یال‌ها (مطابق با الگوی از پیش تعریف‌شده) می‌دانند که از ترکیب زیرگراف‌های کوچک‌تر شبکه با یکدیگر حاصل می‌شود.

به بیان دیگر، یک انجمن، مجموعه‌ای از گره‌ها با تعداد مشخصی یال بین آن‌ها است که از یک توپولوژی واضح و مشخص، مطابق مفاهیم ریاضی گراف شبکه، نتیجه می‌شود.

اساس کار این روش‌ها بر مفهومی ساده به نام کلیک استوار است. منظور از عبارت k-clique در گراف یک شبکه، یک زیرگراف شامل k رأس است که $k\times(k-1)/2$ یال دارد. انتخاب k=3 کوچک‌ترین کلیک غیربدیهی (یک مثلث) را نشان می‌دهد و انتخاب مقادیر بزرگ‌تر برای k، باعث یافتن کلیک‌های بزرگ‌تر و با ساختار متراکم‌تر می‌شود. مجاز بودن تعلق هم‌زمان رئوس به دو یا چند کلیک متمایز، هم‌پوشانی را مدل می‌کند.

روش CPM[7] [۳۷] انجمن‌های هم‌پوشان یک گراف را به کمک کلیک‌های مجاور به دست می‌آورد. دو k-clique مجاور یکدیگر هستند، اگر در k-1 رأس با یکدیگر مشترک باشند (فقط در یک رأس تفاوت داشته باشند). این روش از تعاریف و مفاهیم دیگری نظیر اجتماع، زنجیره، ارتباط و انجمن روی k-cliques استفاده می‌کند. CPM در ابتدا همه کلیک‌ها به اندازه k را می‌یابد و با کمک مفاهیم فوق، انجمن‌ها را روی اجتماع این کلیک‌ها تعریف می‌کند. این روش ساده و متداول، زمان اجرای طولانی دارد و فقط برای گراف‌های کوچک مناسب است.

تلاش‌هایی برای افزایش سرعت CPM انجام شده است. روش EAGLE[8] از ایدهٔ روش‌های انتشار و تلاش برای بهینه‌سازی محلی بهره می‌برد [۳۸]، با این تفاوت که به جای استفاده از گره‌ها به‌عنوان نقاط شروع اولیه، از کلیک‌ها برای این کار استفاده می‌شود. ایدهٔ روش در این است که نواحی چگال یک شبکه، به احتمال زیاد، یک کلیک بزرگ (به‌عنوان هستهٔ یک انجمن) دارند.

یک ماکزیمال کلیک[9] ، کلیکی است که خود زیرمجموعهٔ هیچ کلیک بزرگ‌تری نباشد. این روش، در ابتدا تمام

[6] Inflation
[7] Clique Percolation Method (CPM)
[8] agglomerativE hierarchicAl clusterinG based on maximaL cliquE (EAGLE)
[9] Maximal Clique


www.SID.ir

Archive of SID

ماکزیمال‌کلیک‌های موجود در گراف را می‌یابد و برای هر زوج دلخواه از این ماکزیمال کلیک‌ها، به بررسی میزان شباهت می‌پردازد. این بررسی با یک تابع پیمانه‌ای تعمیم‌یافته انجام می‌شود که از مفهوم هم‌پوشانی حمایت می‌کند. اگر مقدار آن از یک حدآستانه بیشتر بود، آن دو ماکزیمال کلیک را ادغام می‌کند و مراحل را تا رسیدن به یک انجمن ادامه می‌دهد.

روش GCE[1]، با ایده‌ای مشابه روش EAGLE از ماکزیمال کلیک‌ها به‌عنوان نقاط شروع انتشار و بسط انجمن‌ها استفاده می‌کند [۳۹]؛ امّا برخلاف آن روش که از تابع پیمانه‌ای سود می‌برد، خروجی انجمن‌یابی را بر اساس تابع بیان‌شده در رابطه (۶) تعیین می‌کند:

$$f(c) = 1 - \frac{|c_1 \cap c_2|}{\min(|c_1|, |c_2|)} \qquad (۶)$$

که در آن، $|c_1|$ تعداد گره‌های انجمن $c_1$ است.

### ۲-۵- انجمن‌یابی مبتنی بر نزدیکی [2]

گروهی از روش‌ها، بر اساس مفهوم «جهان کوچک[3]» عمل می‌کنند. شبکه‌های جهان کوچک دارای نسبت بالایی از اتصال گره‌ها به یکدیگر بوده، متوسط طول مسیر در آن‌ها (با وجود بزرگ بودن خود شبکه)، عدد کوچکی است.

ایدهٔ کلی در این است که هرچه فاصله میان دو گره کمتر باشد (همسایگان، همسایگان همسایگان و ...)، احتمال تعلق آن‌ها به یک انجمن بیشتر است. در این روش‌ها، استفاده از مکانیزم «گام تصادفی[4]»، کاربردی فراوان دارد که به حرکت یک گام‌زننده از رأس موردنظر به رئوس هم‌جوار با انجام چندین گام تصادفی و دسته‌بندی گره‌هایی که بیشتر در یک گام ظاهر می‌شوند، اشاره دارد.

در روش WalkTrap تلاش می‌شود گام‌زننده در نواحی با چگالی بالا (انجمن‌ها) به دام بیفتد [۴۰]. انتخاب طول گام تصادفی، مسئله‌ای مهم است؛ زیرا باید به اندازهٔ کافی بزرگ باشد تا بتواند اطلاعات مناسب را گردآوری کند. طول آن معمولاً بر اساس توپولوژی شبکه تعیین می‌شود. این روش ناهم‌پوشان، یک تابع را برای فاصله بین دو گره تعریف می‌کند که مقدار آن بر اساس اطلاعات به‌دست‌آمده از گام زدن تصادفی در گراف، محاسبه می‌شود.

روش Infomap [۴۱] از ترکیب تکنیک‌های مبتنی بر تئوری اطلاعات با معلومات ناشی از گام‌های تصادفی سود می‌برد. این روش یافتن بهترین ساختار انجمنی را معادل با کمینه‌سازی کمیت اطلاعات موردنیاز برای بیان گام تصادفی در یک شبکه می‌داند. کشف ساختار گراف با تعدادی گام تصادفی با طول‌های متفاوت و با بررسی احتمال پرش به یک گره تصادفی دیگر (مشابه با ایدهٔ الگوریتم رتبه‌صفحه گوگل در جست‌وجو میان صفحات وب) انجام خواهد شد. برای هر گره انجمن یک نام منحصربه‌فرد انتخاب می‌شود. در این حالت، گره‌های داخلی انجمن (گره‌های پرتکرار) نام کوتاه‌تری دارند. فشرده‌سازی مناسب در دو مرحله توسط کد هافمن انجام خواهد شد که یک مرحلهٔ آن برای تشخیص انجمن‌ها در گراف شبکه و مرحلهٔ دیگر برای تشخیص گره‌های درون یک انجمن است. بهترین حالت انجمن‌یابی، کوتاهترین طول نام‌گذاری (توصیف) را در میان گام‌های تصادفی ممکن خواهد داشت. روش Infomap یک روش پرکاربرد و دقیق است؛ امّا هم‌پوشان نیست.

در روش UEOC[5] که هم‌پوشانی را نیز حمایت می‌کند، برای تخصیص هر گره به یک انجمن، گام‌زننده تصادفی با زنجیرهٔ مارکوف مدل می‌شود [۴۲]. انجمن یک گره، بر اساس احتمال گذر L- گامی یک گام تصادفی مارکوف روی شبکه تبرید[6] متناظر با شبکهٔ گراف اصلی تعیین می‌شود. شبکهٔ تبرید، شبکه‌ای تصادفی است که با استفاده از یک مکانیزم تجمیعی[7]، درجات گره‌ها را حفظ می‌کند. مرتب‌سازی گره‌ها بر اساس احتمال گذر آن‌ها به‌صورت نزولی است. تعیین و استخراج انجمن‌ها به کمک معیار کندوکتانس (توضیح در بخش ۴) انجام می‌شود.

### ۲-۶- انجمن‌یابی مبتنی بر خوشه‌بندی یال‌ها

روش‌های مطرح‌شده در بخش‌های قبلی، انجمن‌ها را مجموعه‌ای از گره‌ها فرض کرده‌اند؛ امّا روش‌هایی وجود دارند که اصالت را به یال‌ها می‌دهند و انجمن را قسمتی از یک شبکه می‌دانند که یال‌های داخلی‌اش بیشتر از یال‌های خارجی آن است.

مکانیزم این روش‌ها، مبتنی بر استفاده از «گراف خط[8]»

[1] Greedy Clique Expansion (GCE)
[2] Closeness
[3] Small World
[4] Random Walk
[5] Unfold and Extract Overlapping Communities (UEOC)
[6] Annealed
[7] Ensembling
[8] Line Graph


www.SID.ir

Archive of SID

یا L(G) است که متناظر با هر گراف G(V,E) به فرم گرافی تعریف می‌شود که هر گره آن نظیر یک یال از گراف اصلی و هر یال آن، بیانگر وجود رأس مشترک بین هر دو یال از گراف اصلی باشد؛ مثلاً گراف خط متناظر با یک گراف ستاره، یک گراف کامل است که به فرم یک کلیک به دست می‌آید و تعداد رئوس آن برابر با تعداد یال‌های گراف ستاره است.

هر گره در گراف اصلی یک گره هم‌پوشان تلقی می‌شود، اگر یال‌های متصل به آن در بیش از یک خوشه از گراف خط نظیر گراف اصلی، عضویت داشته باشند. بنابراین، هر افرازی در گراف خط معادل با یک انجمن‌یابی هم‌پوشان در گراف اصلی است.

روش موسوم به LC[1] برای یافتن زیرگراف‌های هم‌پوشان از گام تصادفی کمک می‌گیرد [۴۳]. این روش به پایش مقدار تابع پیمانه‌ای در حین گام‌زدن تصادفی روی گراف و تولید دنباله‌های یال- یال و یال- رأس- یال می‌پردازد. یک گراف خط وزن‌دار تولید می‌شود تا بر اساس مقادیر حاصل از تابع پیمانه‌ای دربارهٔ افراز مناسب برای یال‌های آن تصمیم‌گیری شود.

روش HLC[2] ، تعلق یک گره به دو یا چندین انجمن را با خوشه‌بندی یال‌ها مدل می‌کند [۴۴]. در این روش، یک معیار شباهت یال تعریف می‌شود تا برای مجموعه‌ای از همسایگان مشترک گره‌ها محاسبه شود.

شباهت یال برای دو یال $e_{ik}$ و $e_{jk}$ که در رأس $k$ مشترک هستند، مطابق رابطه (۷) تعریف می‌شود:

$$S(e_{ik}, e_{jk}) = \frac{|n_+(i) \cap n_+(j)|}{|n_+(i) \cup n_+(j)|} \qquad (۷)$$

که در آن، $n_+(i)$ مجموعه همسایگان گره $i$ است و $e_{ik}$ معرف یال موجود بین گره‌های $i$ و $k$. این معیار معادل با نسبت تعداد همسایگان مشترک میان دو رأس به تعداد کل همسایگان هر دو رأس، تعریف می‌شود.

در انتها، یک روش کلاسیک خوشه‌بندی سلسله‌مراتبی برای ساخت دندوگرام یال اعمال می‌شود تا قطع کردن دندوگرام حاصله در مکان و مقدار مناسب، انجمن‌یابی بر اساس یال‌ها را محقق کند.

### ۲-۷- جدول ویژگی روش‌ها

در بخش‌های ۲-۱ تا ۲-۶ این مقاله، روش‌های مطرح در انجمن‌یابی و نحوهٔ مدل‌کردن هم‌پوشانی توسط آن‌ها در شش دستهٔ نسبتاً مجزّا با توصیفاتی چون استفاده از مفاهیم طیف و معیارهای مرکزگرایی، تعریف یک تابع کیفیت (چگالی)، رفتار یکسان در انتشار یک خاصیت، جست‌وجوی یک ساختار خاص در گراف، نزدیکی گره‌های شبکه به یکدیگر و نهایتاً خوشه‌بندی یال‌ها به جای گره‌ها بررسی شده‌اند. در اینجا از عبارت «نسبتاً مجزا» استفاده شده است؛ چون ایده‌های به‌کاررفته در شکل‌گیری یک انجمن، خود ممکن است دارای اشتراکاتی باشند؛ مثلاً روشی که کلیک را (به جای گره) به‌عنوان هستهٔ اولیه برای انتشار انجمن‌ها مدنظر قرار می‌دهد، هم از مفاهیم بخش ساختار و هم از مفاهیم بخش انتشار استفاده می‌کند.

در جدول ۱ روش‌های انجمن‌یابی مطرح‌شده در ۲-۱ تا ۲-۶، با ذکر ویژگی‌های مهم آن‌ها فهرست شده‌اند. به همراه نام اختصاری هر روش، اطلاعاتی چون مدل مورد استفاده برای تعریف و تشکیل یک انجمن، امکان اعمال روش مربوط روی گراف‌های جهت‌دار و وزن‌دار، نیاز روش به پارامتر ورودی، پیچیدگی زمانی روش، حمایت از مفهوم هم‌پوشانی و نهایتاً سال ارائه روش، بیان شده است.

در بررسی قابلیت روش برای کار با گراف‌های جهت‌دار و وزن‌دار، ویژگی‌های رایج‌ترین نسخه یک روش، مدنظر بوده است. توضیح اینکه، معمولاً تعمیم‌های متعددی برای روش‌های معروف ارائه می‌شود تا بعضی قابلیت‌ها را به نسخهٔ اولیهٔ آن اضافه کند.

عمدتاً رسیدن به این قابلیت‌ها، با پیچیدگی زمانی بالاتر یا کم‌رنگ شدن دیگر مزیت‌های روش همراه است.

به‌عنوان مثال، می‌توان به تعمیم‌های ارائه‌شده روی CPM اشاره کرد تا این روش هم امکان کار با گراف‌های دارای وزن و جهت را داشته باشد. دو تعمیم پیشنهادی شامل روش CPMd [۴۵] که از مفهوم کلیک روی گراف‌های جهت‌دار (کلیک جهت‌دار) و روش CPMw [۴۶] که از مفهوم کلیک روی گراف‌های وزن‌دار (کلیک وزن‌دار) سود می‌برند، هیچ‌گاه به اندازهٔ روش اصلی متداول نشده‌اند.

عدم نیاز به پارامتر ورودی، به معنای سادگی یک روش است و مزیتی برای آن تلقی می‌شود. در روش‌های بدون پارامتر ورودی، فقط لازم است گراف ورودی به‌صورت

[1] Link Communities (LC = LinkCom)

[2] Hierarchical Link Clustering (HLC)


www.SID.ir

Archive of SID

ماتریس مجاورت یا لیست مجاورت به روش مربوط، داده شود. به‌عنوان مثال‌های شاخص از پارامتر ورودی می‌توان به متغیرهای توزیع آماری یا مقدار k در تعیین اندازه کلیک، اشاره کرد.

جدول ۱: مشخصات الگوریتم‌های انجمن‌یابی

| نام روش | مدل | پیچیدگی | جهت‌دار بودن گراف ورودی | وزن‌دار بودن گراف ورودی | نیاز به پارامتر ورودی | هم‌پوشانی | سال |
|---|---|---|---|---|---|---|---|
| Bayesian-NMF | طیفی | $O(cn^2)$ | × | ✓ | دارد | ✓ | ۲۰۱۱ |
| BIG CLAM | طیفی | $O(cn + m)$ | × | ✓ | دارد | ✓ | ۲۰۱۳ |
| GN | مرکزگرایی | $O(m^2n)$ | ✓ | ✓ | دارد | × | ۲۰۰۲ |
| CONGA | مرکزگرایی | Best: $(nlogn)$ Worst: $O(m^3)$ | × | ✓ | دارد | ✓ | ۲۰۰۷ |
| DOCNet | مرکزگرایی | $O(n^2)$ | × | × | ندارد | ✓ | ۲۰۱۴ |
| NLA | مرکزگرایی | $O(n^2)$ | ✓ | ✓ | دارد | ✓ | ۲۰۱۶ |
| CNM | تابع کیفیت | $O(nlog^2n)$ | × | ✓ | ندارد | × | ۲۰۰۴ |
| Louvain | تابع کیفیت | $O(m)$ | × | ✓ | ندارد | × | ۲۰۰۸ |
| CIS | تابع کیفیت | $O(n^2)$ | × | ✓ | دارد | ✓ | ۲۰۰۹ |
| LFM | تابع کیفیت | $O(n^2)$ | × | ✓ | دارد | ✓ | ۲۰۰۹ |
| OSLOM | تابع کیفیت | $O(n^2)$ | ✓ | ✓ | دارد | ✓ | ۲۰۱۱ |
| FOCS | تابع کیفیت | $O(m + n)$ | × | × | دارد | ✓ | ۲۰۱۵ |
| GA-Net | تابع کیفیت | $O(n^2)$ | × | × | دارد | ✓ | ۲۰۰۸ |
| Game | تابع کیفیت | $O(m^2)$ | ✓ | ✓ | دارد | ✓ | ۲۰۱۰ |
| LPA | انتشار | $O(m + n)$ | × | ✓ | ندارد | × | ۲۰۰۷ |
| COPRA | انتشار | $O(mlog(m/n))$ | × | ✓ | ندارد | ✓ | ۲۰۱۰ |
| SLPA | انتشار | $O(Tm)$ | ✓ | ✓ | دارد | ✓ | ۲۰۱۲ |
| DEMON | انتشار | $O(nK^{3-\alpha})$ | × | × | ندارد | ✓ | ۲۰۱۲ |
| MMSB | انتشار | $O(\|c\|.n)$ | ✓ | × | دارد | ✓ | ۲۰۰۷ |
| MOSES | انتشار | $O(n^2)$ | × | × | ندارد | ✓ | ۲۰۱۰ |
| NISE | انتشار | $O(m + n)$ | × | ✓ | دارد | ✓ | ۲۰۱۶ |
| CPM | ساختار | $O(m^{\log m/10})$ | × | × | دارد | ✓ | ۲۰۰۵ |
| EAGLE | ساختار | $O(n^2+(h+n)p)$ | ✓ | ✓ | دارد | ✓ | ۲۰۰۹ |
| GCE | ساختار | $O(mh)$ | ✓ | ✓ | دارد | ✓ | ۲۰۱۰ |
| WalkTrap | نزدیکی | $O(mn^2)$ | × | ✓ | دارد | × | ۲۰۰۶ |
| Infomap | نزدیکی | $O(nlogn)$ | ✓ | ✓ | دارد | × | ۲۰۰۸ |
| UEOC | نزدیکی | $O(ln^2)$ | × | × | دارد | ✓ | ۲۰۱۱ |
| LinkCom | یال | $O(m.\|c\|.logn)$ | × | ✓ | دارد | ✓ | ۲۰۰۹ |
| HLC | یال | $O(nK^2)$ | × | ✓ | دارد | ✓ | ۲۰۱۰ |

(n= تعداد رئوس، m= تعداد یال‌ها، K= بیشینه درجات رئوس، α= ضریب توانی، $|c|$= تعداد انجمن‌ها، T= تعداد دفعات اجرای الگوریتم، h= تعداد ماکزیمال‌کلیک‌های یک گراف، p= تعداد زوج‌های ماکزیمال‌کلیک مجاور یکدیگر، l= طول گام تصادفی)


www.SID.ir

Archive of SID

پیچیدگی زمانی هر روش برای بدترین حالت آن ذکر شده است. در بیشتر موارد، پیچیدگی زمانی روش‌ها برای حالت متوسط و بدترین حالت آن، یکسان است. یادآوری می‌شود که مقدار m برای گراف‌های کم‌تراکم، متناسب با n و برای گراف‌های متراکم، متناسب با $n^2$ است.

با نگاه به جدول ۱ که بسیاری از مطالب مطرح‌شده در ۲-۱ تا ۲-۶ را تجمیع و خلاصه کرده است، یک نکته مهم مشخص می‌شود. برای همین خاصیت‌های مطرح‌شده نیز هیچ روش انجمن‌یابی وجود ندارد که هم از هم‌پوشانی روی گراف‌های جهت‌دار و وزن‌دار حمایت کند، هم پارامتر ورودی نداشته باشد و هم پیچیدگی زمانی آن مطلوب باشد. هریک از روش‌ها با توجه به مکانیزم خود، دارای قوّت‌ها و ضعف‌های متعددی هستند. در بخش بعدی، به بررسی کمّی این قوّت‌ها و ضعف‌ها می‌پردازیم.

## ۳- مقایسهٔ عملکرد روش‌ها

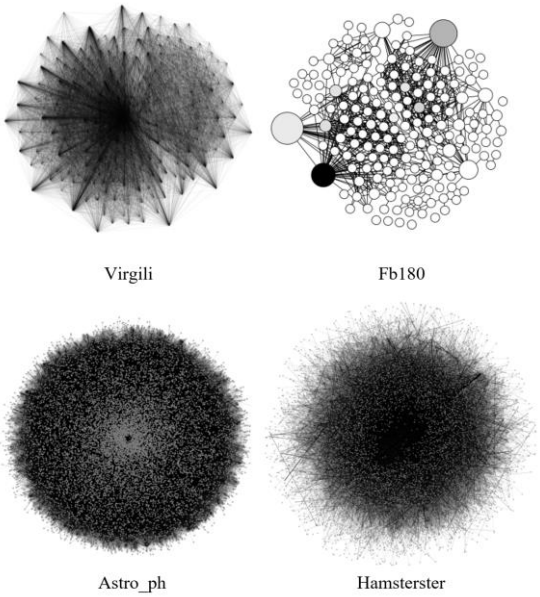
در این بخش، به بررسی عملکرد تعدادی از روش‌های رایج در انجمن‌یابی می‌پردازیم. بخش ۳-۱، به معرفی چهار دادگان[1] انتخابی و بیان ویژگی‌های آن‌ها اختصاص دارد. در بخش ۳-۲ چند معیار ارزیابی کارایی معرفی می‌شود تا در برآورد میزان مطلوب بودن انجمن‌یابی استفاده گردد. بخش ۳-۳، به تحلیل و بررسی نتایج حاصل از معیارهای ارزیابی می‌پردازد.

### ۳-۱- مشخصات دادگان‌ها

مطالعات این مقاله روی چهار دادگان دنیای واقعی انجام می‌شود که در مقایسه با یکدیگر، از تنوع خوبی در اندازه، چگالی و دیگر مشخصات گراف برخوردارند.

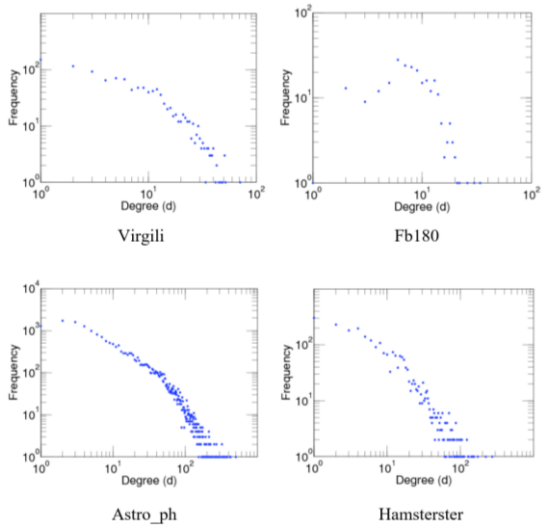
اولین دادگان با نام Fb180، گراف دوستان یک کاربر شبکه فیس‌بوک بوده که با نرم‌افزار Netvizz از این شبکه استخراج شده است. دادگان دوم، یک شبکه ارتباطی میان گروهی از دانشجویان دانشگاه Virgili در کشور اسپانیا و دادگان سوم مربوط به پیوندهای بین کاربران در سایت hamsterster.com است. هر دو دادگان از آرشیو پایگاه KONECT [۴۷] انتخاب شده‌اند. نهایتاً دادگان چهارم به نام Astro_ph، همکاری‌های علمی گروهی از اختر فیزیک‌دانان (ثبت‌شده در سایت arXiv) را نشان می‌دهد. این دادگان از پایگاه SNAP [۴۸] انتخاب شده است. در انتخاب دادگان‌ها، هم از دادگان تولیدشده توسط نویسندگان مقاله و هم از دادگان استاندارد و پرکاربرد، استفاده کرده‌ایم.

جدول ۲ مشخصات اولیهٔ هریک از دادگان‌ها و مقادیر پارامترهای مهم گراف متناظر آن‌ها را بیان می‌کند. شکل (۴) یک تجسم از این دادگان‌ها را نمایش می‌دهد. رسم شکل به کمک نرم‌افزار Gephi [۴۹] انجام شده است.




شکل ۴: گراف‌های متناظر با چهار دادگان

با توجه به درجهٔ هر گره در گراف متناظر با هر دادگان، منحنی توزیع درجات برای دادگان‌ها به دست می‌آید. این منحنی‌ها در شکل (۵) قابل مشاهده هستند.




شکل ۵: توزیع درجات گراف در چهار دادگان

[1] Dataset


www.SID.ir

Archive of SID

جدول ۲: مشخصات چهار دادگان

| دادگان | n | M | متوسط درجه یک گره | بیشینه درجه یک گره | تعداد مؤلّفه‌های همبند | nGCC | mGCC | چگالی گراف | قطر گراف | APL |
|---|---|---|---|---|---|---|---|---|---|---|
| Fb180 | ۱۸۰ | ۷۵۴ | ۸,۳۷ | ۵۴ | ۱۳ | ۱۵۲ | ۷۲۷ | ۰,۰۴۶۸۰۳ | ۸ | ۳,۲۴ |
| Virgili | ۱۱۳۳ | ۵۴۵۱ | ۹,۶۲ | ۷۱ | ۱ | ۱۱۳۳ | ۵۴۵۱ | ۰,۰۰۸۵۰۰ | ۸ | ۳,۶۵ |
| Hamsterster | ۲۴۲۶ | ۱۶۶۳۱ | ۱۳,۷۱ | ۲۷۳ | ۶۹ | ۲۰۰۰ | ۱۵۷۲۳ | ۰,۰۰۵۶۵۳ | ۱۰ | ۳,۶۷ |
| Astro_ph | ۱۸۷۷۱ | ۱۹۸۰۵۰ | ۲۱,۱۰ | ۵۰۴ | ۵۳ | ۱۷۹۰۳ | ۱۸۹۷۵۷ | ۰,۰۰۱۱۲۴ | ۱۴ | ۴,۱۷ |

کسب اطلاعات اولیه دربارهٔ شبکه‌های کوچک، با نگاه به گراف آن‌ها امکان‌پذیر است؛ مثلاً برای دادگان اول در شکل (۴)، اندازهٔ هر گره متناسب با بینابینی آن و شدّت رنگ هر گره متناسب با درجهٔ آن انتخاب شده است. مشخص است که برای این دادگان، گره دارای بیشترین درجه، الزاماً همان گره دارای بیشترین بینابینی نیست.

همچنین در شکل (۵) با وجود تفاوت در تعداد رئوس و یال‌ها، این نکته مشهود است که دادگان‌ها توزیع توانی تقریباً مشابه با یکدیگر دارند و جزء شبکه‌های مقیاس‌آزاد محسوب می‌شوند.

توپولوژی یک گراف وابسته به ماهیت و نحوه تشکیل آن بر اساس ارتباطات تعریف‌شده میان موجودیت‌هایش است. بنابراین ممکن است یک گراف بزرگ، یکپارچه و همبند[1] بوده و فقط از یک (یا چند مؤلّفه معدود) تشکیل شده باشد و در مقابل، گرافی با $m$ و $n$ کمتر، تعداد مؤلّفه‌های همبند بیشتری داشته باشد. برای چهار دادگان انتخابی، تعداد مؤلّفه‌های همبند از ۱ تا ۶۹ متغیر است.

در گراف‌های ناهمبند، بسیاری از پردازش‌های گراف روی «بزرگ‌ترین مؤلّفه همبند[2]» انجام می‌شود. نظر به اهمیت این مؤلّفه، تعداد رئوس ($n_{GCC}$) و تعداد یال‌های آن ($m_{GCC}$) برای هر یک از دادگان‌ها در جدول ۲ ذکر شده است. برای یک گراف با $n$ رأس و $m$ یال، چگالی یا میزان تراکم گراف که بیانگر نسبت یال‌های موجود به یال‌های ممکن است، با $2m/n(n-1)$ بیان می‌شود.

طول مسیر بین دو گره، با حداقل تعداد یال‌های ممکن میان آن دو گره مشخص می‌شود. مفهوم «قطر گراف»، بیانگر مقدار بیشینه در طول مسیر بین دو گره از گراف است و مفهوم APL[3] به متوسط‌گیری روی تمام این طول مسیرها اشاره دارد.

در گراف‌های با چند مؤلّفه همبند، مقادیر قطر و APL روی بزرگ‌ترین مؤلّفهٔ همبند گراف تعیین می‌شوند. با توجه به مقادیر APL در جدول ۲، دادگان‌های انتخابی جزء شبکه‌های جهان کوچک هستند.

### ۳-۲- معرفی معیارهای ارزیابی کارایی

معیارهای زیادی در بررسی عملکرد روش‌های انجمن‌یابی وجود دارند. برخی از معیارها فقط در حالت ناهم‌پوشان و برخی فقط در حالت هم‌پوشان استفاده می‌شوند. گروهی از معیارها، قابلیت استفاده در هر دو حالت را دارند.

یک معیار ساده، تعداد انجمن‌های یافته‌شده توسط روش انجمن‌یابی است که با $|c|$ بیان می‌شود. به غیر از مواردی که تعداد انجمن‌ها به‌عنوان یک پارامتر ورودی مسئله داده می‌شود، یافتن مقدار بهینه (یا مناسب) برای $|c|$ به‌شدت به تعریف مسئله و توپولوژی گراف وابسته است.

معیار Flake-ODF [۵۰] برآورد می‌کند که چه کسری از گره‌های انجمن $S$، دارای تعداد یال‌های داخلی کمتر از یال‌های مرزی هستند (یال‌های داخلی آن گره از نصف درجه آن گره کمتر است). حالت مطلوب، در کمتر شدن مقدار این معیار است. این کاهش به معنای وجود شباهت بیشتر در ساختار گره‌های در حال بررسی با ساختار یک انجمن است. این معیار در رابطه (۸) برای انجمن $S$ با $n_S$ گره، تعریف شده است:

$$\text{Flake}-\text{ODF}(S) = \frac{\left|\{u : u \in S, |\{(u,v) : v \in S\}| < \deg(u)/2\}\right|}{n_S} \qquad (۸)$$

منظور از $\deg(u)$، درجه رأس $u$ است. مقدار این معیار برای کل گراف نیز قابل محاسبه است (با متوسط‌گیری روی

[1] Connected
[2] Greatest (Giant) Connected Component (GCC)
[3] Average Path Length (APL)


www.SID.ir

Archive of SID

همهٔ انجمن‌ها).

معیار کندوکتانس [۵۰] نقش یال‌های مرزی یک انجمن (حاصل از الگوریتم انجمن‌یابی) را نشان می‌دهد. مطابق رابطه (۹)، تعداد یال‌های داخلی انجمن S ( تعلق هر دو رأس به S) با $m_S$ بیان می‌شود:

$$m_S = \left|\{(u,v) : u \in S, v \in S\}\right| \qquad (۹)$$

تعداد یال‌های مرزی انجمن S (یالی که فقط یک رأس آن به انجمن S تعلق دارد) مطابق رابطه (۱۰) با $c_S$ نشان داده شده است:

$$c_S = \left|\{(u,v) : u \in S, v \notin S\}\right| \qquad (۱۰)$$

با توجه به روابط (۹) و (۱۰)، کندوکتانس مطابق رابطه (۱۱) محاسبه می‌شود:

$$cond(S) = \frac{c_S}{2m_S + c_S} \qquad (۱۱)$$

این معیار برای کل گراف هم به‌سادگی با متوسط‌گیری روی تمام انجمن‌ها، قابل محاسبه است. کمتر بودن مقدار این معیار، معادلِ یافتن ساختار انجمنی قوی‌تر و عملکرد مطلوب‌تر در روش مورد استفاده است.

معیار «نسبت هم‌پوشانی[1] » [۸] مطابق رابطه (۱۲) روی گراف تعریف می‌شود:

$$ov - ratio = \frac{1}{n}\sum_{i=1}^{n}\left|u_i \in S, S \in G\right| \qquad (۱۲)$$

برای هر رأس $u_i$، تعداد انجمن‌هایی که به آن‌ها تعلق دارد، مشخص می‌شود و متوسط‌گیری روی مقادیر رئوس گراف انجام خواهد شد.

جدول ۳: عملکرد الگوریتم‌های انجمن‌یابی روی چهار دادگان

| | | GN | CONGA | LFM | LPA | COPRA | CPM | GCE | Infomap | HLC |
|---|---|---|---|---|---|---|---|---|---|---|
| Fb180 | \|c\| | ۱۸ | **۳۰** | ۱۱ | ۱۹ | ۲۷ | ۸ | ۹ | ۱۸ | ۲۹ |
| | Q | ۰,۴۴ | ۰,۴۷ | ۰,۴۸ | ۰,۴۴ | **۰,۴۹** | ۰,۴۰ | ۰,۴۱ | ۰,۴۲ | ۰,۴۴ |
| | Flake-ODF | ۰,۱۲ | ۰,۲۵ | ۰,۱۰ | ۰,۰۵ | ۰,۲۷ | **۰,۰۴** | **۰,۰۴** | ۰,۱۱ | ۰,۱۰ |
| | Conductance | **۰,۰۳** | ۰,۱۷ | ۰,۰۸ | ۰,۰۴ | ۰,۱۴ | ۰,۰۶ | ۰,۰۶ | ۰,۰۴ | ۰,۰۹ |
| | Ov-ratio | ۱ | ۲,۴۱ | ۱,۲۳ | ۱ | **۳,۱۴** | ۰,۷۱ | ۰,۷۳ | ۱ | ۲,۷۹ |
| Virgili | \|c\| | ۴۹ | ۱۱۷ | ۶۱ | ۵۲ | ۷۸ | ۴۰ | ۵۵ | ۵۸ | **۱۳۰** |
| | Q | ۰,۳۸ | ۰,۳۹ | ۰,۴۰ | ۰,۳۵ | **۰,۴۱** | ۰,۳۰ | ۰,۳۲ | ۰,۳۴ | ۰,۳۷ |
| | Flake-ODF | ۰,۱۱ | ۰,۲۰ | ۰,۱۱ | ۰,۰۷ | ۰,۲۱ | **۰,۰۵** | ۰,۰۶ | ۰,۰۸ | ۰,۱۱ |
| | Conductance | **۰,۰۴** | ۰,۱۹ | ۰,۱۰ | ۰,۰۶ | ۰,۱۶ | ۰,۰۸ | ۰,۰۸ | ۰,۰۷ | ۰,۱۱ |
| | Ov-ratio | ۱ | ۲,۰۴ | ۱,۳۰ | ۱ | ۲,۲۸ | ۰,۷۸ | ۰,۸۲ | ۱ | **۲,۳۳** |
| Hamsterster | \|c\| | ۵۹۰ | ۱۰۴۷ | ۹۲۱ | ۵۸۲ | ۷۷۰ | ۴۷۳ | ۵۱۱ | ۷۲۶ | **۱۱۳۸** |
| | Q | ۰,۲۷ | ۰,۳۰ | **۰,۳۲** | ۰,۲۹ | ۰,۳۱ | ۰,۱۸ | ۰,۲۱ | ۰,۲۵ | ۰,۲۶ |
| | Flake-ODF | ۰,۱۴ | ۰,۲۳ | ۰,۱۶ | ۰,۱۶ | ۰,۲۰ | **۰,۰۹** | ۰,۱۰ | ۰,۱۱ | ۰,۱۴ |
| | Conductance | **۰,۰۷** | ۰,۲۴ | ۰,۱۳ | ۰,۱۰ | ۰,۲۰ | ۰,۱۲ | ۰,۱۳ | ۰,۰۹ | ۰,۱۵ |
| | Ov-ratio | ۱ | ۱,۸۵ | ۱,۴۵ | ۱ | ۲,۳۱ | ۰,۶۳ | ۰,۷۱ | ۱ | **۲,۳۵** |
| Astro_ph | \|c\| | ۱۲۹۱ | ۴۰۱۷ | ۱۵۷۷ | ۹۵۱ | ۱۶۵۴ | ۷۹۲ | ۸۳۹ | ۱۰۸۵ | **۴۱۸۲** |
| | Q | ۰,۱۳ | ۰,۲۰ | **۰,۲۲** | ۰,۱۵ | **۰,۲۲** | ۰,۱۱ | ۰,۱۳ | ۰,۱۴ | ۰,۱۶ |
| | Flake-ODF | ۰,۱۵ | ۰,۲۳ | ۰,۱۳ | ۰,۱۴ | ۰,۲۳ | **۰,۰۷** | ۰,۰۹ | ۰,۱۳ | ۰,۱۵ |
| | Conductance | **۰,۰۵** | ۰,۲۰ | ۰,۱۴ | ۰,۰۷ | ۰,۱۷ | ۰,۰۹ | ۰,۰۹ | ۰,۰۷ | ۰,۱۱ |
| | Ov-ratio | ۱ | ۱,۵۳ | ۱,۵۸ | ۱ | ۱,۷۴ | ۰,۶۹ | ۰,۷۷ | ۱ | **۲,۰۹** |

[1] Overlap Ratio


www.SID.ir

Archive of SID

### ۳-۳- بررسی نتایج

این بخش به توضیحات مربوط به پیاده‌سازی چند روش انجمن‌یابی روی دادگان‌های معرفی‌شده در ۳-۱ و مقایسهٔ عملکرد روش‌ها با یکدیگر اختصاص دارد.

برای یکسان‌سازی در مقایسهٔ روش‌ها، پیاده‌سازی‌ها روی حالت بدون وزن و غیرجهت‌دار گراف دادگان‌های انتخابی انجام شده است. مکانیزم نُه روش (شش روش هم‌پوشان و سه روش ناهم‌پوشان) در پیاده‌سازی استفاده شد تا برای هرکدام از دسته‌های مطرح‌شده در ۲-۱ تا ۲-۶، عملکرد حداقل یک روش مورد ارزیابی قرار گرفته باشد. در جدول ۳ نتایج حاصل از این روش‌های انجمن‌یابی، توسط چند معیار ارزیابی کارایی بیان شده است.

پیاده‌سازی COPRA[1] و پیاده‌سازی CPM[2] با زبان جاوا و پیاده‌سازی LFM[3] با زبان C++ (بهره‌گیری از کدهای منتشرشده توسط نویسندگان مقالات) انجام شده است.

پیاده‌سازی الگوریتم دیگر روش‌ها، با زبان پایتون (استفاده از ماژول Networkx آن) انجام شد. البته با توجه به غیرقطعی‌بودن روش‌های LPA و COPRA، نتیجهٔ نهایی آن‌ها از متوسط‌گیری روی ده بار اجرا محاسبه شده است. مقدار k در روش‌های ساختارگرای CPM و GCE برابر با پنج در نظر گرفته شده و پارامتر $\gamma$ رابطه (۵) روش LFM برابر با ۰٫۸ انتخاب شده است. ضمناً به دلیل ناهمگونی در پلتفرم‌ها و زبان‌های پیاده‌سازی روش‌ها، زمان اجرا را در مقایسهٔ عملکرد روش‌ها لحاظ نکرده‌ایم.

در جدول ۳، بهترین نتیجهٔ به‌دست‌آمده برای هر معیار ارزیابی کارایی، به‌صورت برجسته و سایه‌دار مشخص شده است. در رابطه با $|c|$ باید گفت که افزایش یا کاهش تعداد انجمن‌های یافته‌شده در مقایسه بین روش‌ها، الزاماً بیانگر مزیتی خاص نیست و به سطح تفکیک[4] تعریف‌شده در مطالعهٔ گراف یک شبکه، مربوط می‌شود.

در حالت کلی، انتخاب تعداد انجمن‌ها بین ۱ (انتخاب خود گراف به‌عنوان انجمن) و n (انتخاب هر رأس به‌عنوان انجمن) امکان‌پذیر است. مشخص است که هیچ یک از این دو حالت، اطلاعات خاصی را بیان نمی‌کند.

بنا بر ماهیت عملکرد خود، روش‌های HLC و CONGA عمدتاً بیشترین تعداد انجمن‌ها را یافته‌اند. کمترین مقادیر توسط روش‌های مبتنی بر ساختار نظیر CPM و GCE حاصل شده است. روش‌های ساختاری به دنبال هسته‌های چگال (کلیک‌ها) هستند تا بر پایه آن‌ها، انجمن‌ها را بیابند. این روش‌ها، رئوس با درجه اندک را در انجمن‌یابی شرکت نمی‌دهند. بنابراین می‌توان برای این روش‌ها انتظار داشت که هرچه گراف مورد بررسی، چگال‌تر (درهم‌تنیده‌تر) شود، تعداد انجمن‌های آن کاهش یابد.

برای محاسبهٔ تابع پیمانه‌ای، در روش‌های ناهم‌پوشان از رابطه (۱) و در روش‌های هم‌پوشان از رابطه (۳) استفاده کرده‌ایم. هم‌پوشانی عمدتاً باعث افزایش در مقدار تابع پیمانه‌ای می‌شود؛ زیرا ترم‌های بیشتری را در محاسبهٔ آن دخالت می‌دهد. بیشترین مقادیر این معیار برای روش‌های LFM و COPRA و کمترین مقادیر برای CPM حاصل شده است.

مقادیر مندرج در جدول ۳ برای معیارFlake-ODF، با متوسط‌گیری از مقادیر همه انجمن‌ها محاسبه شده است. مطابق جدول ۳، بهترین نتیجه توسط روش‌های CPM و GCE به دست آمده است. استفاده از ساختار کلیک که یک گراف کامل است، سبب شده گره‌های انجمن‌های حاصل از این روش هم‌پوشان، دارای پیوندهای داخلی قوی باشند. این گره‌ها عمدتاً از تعداد بیشتری یال داخلی نسبت به یال مرزی برخوردارند. در مقابل، دو روش هم‌پوشانی COPRA و CONGA نتیجهٔ خوبی برای این معیار ندارند؛ زیرا به دلیل نحوهٔ عملکردشان، انجمن‌های زیادی تولید می‌کنند که این امر، سبب افزایش تعداد یال‌های مرزی می‌شود.

مطابق جدول ۳، بهترین نتایج برای معیار کندوکتانس توسط روش ناهم‌پوشان GN حاصل شده است؛ چون بنای کار این روش بر حذف یال‌های فی‌مابین است. می‌توان گفت روش‌های مبتنی بر مرکزگرایی روی دادگان‌هایی که به‌خوبی قابل تفکیک باشند، نتایج مطلوبی می‌گیرند؛ چون بنای کار را بر حذف یال‌ها و نواحی کم‌تراکم می‌گذارند. البته این روش‌ها روی گراف‌های چگال و درهم‌تنیده، به‌خوبی عمل نمی‌کنند.

در انجمن‌یابی، رأس با درجهٔ صفر (رأس ایزوله) به‌عنوان انجمن مستقل در نظر گرفته شده و رأس با درجهٔ یک، به انجمن گره همسایه‌اش اختصاص می‌یابد. بنابراین در روش‌های ناهم‌پوشان، مقدار معیار نسبت هم‌پوشانی برابر با

[1] http://www.cs.bris.ac.uk/~steve/networks/
[2] http://www.cfinder.org
[3] http://sites.google.com/site/andrealancichinetti/software
[4] Granularity


www.SID.ir

Archive of SID

۱ خواهد بود؛ زیرا هر گره فقط و فقط به یک انجمن تعلق دارد.

افزایش مقدار این معیار در روش‌های هم‌پوشان، بیانگر انعطاف روش در تخصیص رئوس به انجمن‌های متعدد است. عمدتاً بیشترین مقدار نسبت هم‌پوشانی برای روش HLC حاصل شده است. شایان ذکر است این معیار برای CPM و GCE، مقداری کمتر از ۱ را نتیجه می‌دهد؛ زیرا این روش‌ها، گره‌های با درجه اندک (کمتر از k-1) را جزء هیچ یک از انجمن‌ها دسته‌بندی نمی‌کنند. مقادیر جدول ۳، برای k=5 به دست آمده‌اند؛ یعنی رئوس با درجه کمتر از ۴ بررسی نمی‌شوند. اگر مقدار بزرگ‌تری انتخاب شود، تعداد انجمن‌ها کاهش خواهد یافت.

انتخاب مقادیر کوچک برای k، افزایش تعداد انجمن‌ها را سبب شده، انتخاب مقادیر خیلی بزرگ برای k، فقط انجمن‌های بسیار چگال را نشان می‌دهد. روش CPM خصوصاً وقتی که گراف یک شبکه دارای تعداد زیادی مؤلّفه کوچک باشد، عملکرد خوبی ندارد.

روش‌های infomap و LPA میانگین عملکرد خوبی دارند، هرچند در هیچ یک از معیارها اول نیستند. ضمناً برخلاف روش‌های ساختاری که در گراف‌های چگال، عملکرد خوبی دارند، روش‌های مبتنی بر یال، نواحی هم‌پوشان متراکم را به‌خوبی تشخیص نداده، ممکن است ناحیهٔ هم‌پوشان متراکم میان دو انجمن متمایز را عملاً یک انجمن مستقل در نظر بگیرند.

## ۴- جمع‌بندی و نتیجه‌گیری

در این مقاله، تعدادی از روش‌های شاخص انجمن‌یابی با توجه به رویکرد روش به تعریف و نحوه شکل‌گیری انجمن و نیز توجه به مفهوم هم‌پوشانی، بررسی شدند.

مدل‌سازی مفهوم هم‌پوشانی (امکان تعلق هم‌زمان اعضا به دو یا چند انجمن) در شش عنوان شامل روش‌های طیفی و مرکزگرایی، تابع کیفیت، انتشار برچسب، ساختار گراف، نزدیکی رئوس به یکدیگر و دسته‌بندی یال‌ها، طبقه‌بندی و بررسی شده‌اند.

استفاده از ایده‌های خلاقانه و متنوع این روش‌ها نظیر کار با معیارهای مرکزگرایی در تفکیک اجزای گراف، عملکرد تابع چگالی، نحوه انتشار اطلاعات، سهولت دسترسی اعضا به یکدیگر و توجه به الگوهای خاص در ساختار گراف، مورد مطالعه و بررسی قرار گرفت.

مقادیر به‌دست‌آمده برای معیارهای ارزیابی کارایی روش‌ها روی مجموعه دادگان انتخابی، نشان می‌دهد که هیچ روشی بهترین مقادیر را برای تمام معیارهای ذکرشده نتیجه نمی‌دهد. هر روش با توجه به مکانیزم عملکرد خود، به تعیین انجمن‌ها می‌پردازد و ممکن است روی شبکه‌ای خاص (با توجه به ویژگی‌ها و توپولوژی آن) نتیجه بهتری حاصل کند. بنابراین هیچ روشی نمی‌تواند خود را به‌عنوان بهترین و جامع‌ترین مکانیزم انجمن‌یابی (خصوصاً برای شبکه‌های دنیای واقعی) مطرح کند. پیشنهاد می‌شود برای هر شبکهٔ خاص، انجمن‌یابی توسط دو یا چند روش انجام شود و نتایج حاصل با هم مقایسه گردد تا بهترین روش با توجه به ویژگی‌های آن شبکه انتخاب شود.


www.SID.ir

Archive of SID

## ۵- مراجع


[1] F. Huang, X. Li, S. Zhang and J. Zhang, "Overlapping Community Detection for Multimedia Social Networks", IEEE Transactions on Multimedia, Vol. 92, 2017, pp. 1-3.

[2] A.L. Barabási and R. Albert, "Emergence of scaling in random networks", Science, Vol. 286, 1999, pp. 509-512.

[3] A.L. Barabási and E. Bonabeau, "Scale-free Networks", Scientific American, Vol. 288, 2003, pp. 50-59.

[4] M. Plantié and M. Crampes, "Survey on social community detection", Social media retrieval (Springer), 2013, pp. 65-85.

[5] S. Yang, X. Yang, C. Zhang and E. Spyrou, "Using social network theory for modeling human mobility", IEEE network, Vol. 24, 2010, pp. 6-13.

[6] S. Fortunato, "Community detection in graphs, Physics reports", Vol. 486, 2010, pp. 75-174.

[7] S. Harenberg, G. Bello, L. Gjeltema, S. Ranshous, J. Harlalka and R. Seay, "Community detection in large-scale networks: a survey and empirical evaluation", Wiley Interdisciplinary Reviews: Computational Statistics, Vol. 6, 2014, pp. 426-439.

[8] M. Coscia, F. Giannotti and D. Pedreschi, "A classification for community discovery methods in complex networks", Statistical Analysis and Data Mining, Vol. 4, 2011, pp. 512-546.

[9] I. Psorakis, S. Roberts, M. Ebden and B. Sheldon, "Overlapping community detection using Bayesian non-negative matrix factorization", Physical Review E, Vol. 83. No. 6, 2011, p. 066114.

[10] J.Yang and J. Leskovec, "Overlapping community detection at scale: a nonnegative matrix factorization approach", in Proceedings of the sixth ACM international conference on Web search and data mining, 2013, pp. 587-596.

[11] M. Girvan and M.E. Newman, "Community structure in social and biological networks", Proceedings of the national academy of sciences, Vol. 99, 2002, pp. 7821-7826.

[12] S. Gregory, "An algorithm to find overlapping community structure in networks", in Proceedings of the European Conference on Principles of Data Mining and Knowledge Discovery, 2007, pp. 91-102.

[13] D. Rhouma and L. Romdhane, "An efficient algorithm for community mining with overlap in social networks", Expert Systems with Applications, Vol. 41, 2014, pp. 4309-4321.

[14] W. Zhi-Xiao, L. Ze-chao, Xiao-fang and T. Jin-hui, "Overlapping community detection based on node location analysis", Knowledge-Based Systems, Vol. 105, 2016, pp. 225-235.

[15] A. Clauset, M.E. Newman and C. Moore, "Finding community structure in very large networks", Physical review E, Vol. 70, 2004, p. 066111.

[16] V.D. Blondel, J. Guillaume, R. Lambiotte and E. Lefebvre, "Fast unfolding of communities in large networks", Journal of statistical mechanics: theory and experiment, 2008, p. 10008.

[17] M. Chen, K. Kuzmin and B.K. Szymanski, "Extension of Modularity Density for overlapping community structure", in Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2014, pp. 856-863.

[18] S. Kelley, "The existence and discovery of overlapping communities in large-scale networks", PhD Thesis, Rensselaer Polytechnic Institute, NY, 2009.

[19] A. Lancichinetti and S. Fortunato, "Community detection algorithms: a comparative analysis", Physical review E, Vol. 80, 2009, p. 056117.

[20] A. Lancichinetti, F. Radicchi, J.J. Ramasco and S. Fortunato, "Finding statistically significant communities in networks", PloS one, Vol. 6, 2011, p. e18961.

[21] S. Bandyopadhyay, G. Chowdhary and D. Sengupta, "FOCS: Fast Overlapped Community Search", IEEE Transactions on Knowledge and Data Engineering, Vol. 27, 2015, pp. 2974-2985.

[۲۲] ح. شریف‌زاده و ن. امجدی، (۱۳۹۳)، «مروری بر انواع الگوریتم‌های فراکاوشی در بهینه‌سازی»، مجله مدل‌سازی در مهندسی، دوره ۱۲، شماره ۳۸، ۱۳۹۳، صفحهٔ ۴۳-۲۷.

[23] C. Pizzuti, "GA-Net: A Genetic Algorithm for Community Detection in Social Networks", Parallel Problem Solving from Nature, PPSN (Springer), 2008, pp. 1081-1090.

[24] C. Pizzuti, "A Multiobjective Genetic Algorithm to Find Communities in Complex Networks", IEEE Transactions on Evolutionary Computation, Vol. 16, 2012, pp. 418-430.

[۲۵] ع. نعیمی‌صدیق، س.ک. چهارسوقی و م. شیخ‌محمدی، «طراحی مدل هماهنگی در زنجیره تأمین رقابتی با استفاده از رویکرد نظریه بازی با همکاری و بدون همکاری»، مجله مدل‌سازی در مهندسی، دوره ۱۰، شماره ۲۹، ۱۳۹۱، صفحهٔ ۳۱-۱۹.

[26] W. Chen, Z. Liu, X. Sun and Y. Wang, "A game-theoretic framework to identify overlapping communities



www.SID.ir

Archive of SID

in social networks", Data Mining and Knowledge Discovery, Vol. 21, 2010, pp. 224-240.

[۲۷] ا. گلکار و م. کائدی، «ارائه مدلی برای تخمین میزان برون‌گرایی اعضای شبکه اجتماعی با استفاده از اطلاعات ساختار گراف»، مجله مدل‌سازی در مهندسی، دوره ۱۳، شماره ۴۳، ۱۳۹۴، صفحهٔ ۱۰۶-۹۱.

[۲۸] ر. احمدی و س. شیخ احمدی، «پیش‌بینی افراد خبره در شبکه‌های اجتماعی آنلاین»، هشتمین کنفرانس بین‌المللی فناوری اطلاعات و دانش، دانشگاه بوعلی سینا، همدان، ۱۳۹۵.

[۲۹] ع.ا. طالعی‌زاده و ز. چراغی، (۱۳۹۴)، «قیمت‌گذاری و بازاریابی در یک زنجیره تأمین دوسطحی تحت چهار رویکرد نظریه بازی‌ها»، مجله مدل‌سازی در مهندسی، دوره ۱۳، شماره ۴۲، ۱۳۹۴، صفحهٔ ۱۴۹-۱۳۵.

[30] U.N. Raghavan, R. Albert and S. Kumara, "Near linear time algorithm to detect community structures in large-scale networks", Physical review E, Vol. 76, 2007, p. 036106.

[31] S. Gregory, "Finding overlapping communities in networks by label propagation", New Journal of Physics, Vol. 12, 2010, p. 103018.

[32] J. Xie and B.K. Szymanski, “Towards linear time overlapping community detection in social networks”, in Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining, 2012, pp. 25-36.

[33] M. Coscia, G. Rossetti, F. Giannotti and D. Pedreschi, “Demon: a local-first discovery method for overlapping communities”, in Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, 2012, pp. 615-623.

[34] E. Airoldi, D. Blei, S. Fienberg and E. Xing, "Mixed membership stochastic blockmodels", Journal of Machine Learning Research, Vol. 9, 2008, pp. 1981-2014.

[35] A. McDaid and N. Hurley, “Detecting highly overlapping communities with model-based overlapping seed expansion”, in Proceedings of the IEEE/ACM International Conference on Social Networks Analysis and Mining (ASONAM), 2010, pp. 112-119.

[36] J. Whang, D. Gleich and I. Dhillon, "Overlapping Community Detection Using Neighborhood-Inflated Seed Expansion", IEEE Transactions on Knowledge and Data Engineering, Vol. 28, 2016, pp. 1272-1284.

[37] G. Palla, I. Derényi, I. Farkas and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society", Nature, Vol. 435, 2005, pp. 814-818.

[38] H. Shen, X. Cheng, K. Cai and M. Hu, "Detect overlapping and hierarchical community structure in networks", Physica A: Statistical Mechanics and its Applications, Vol. 388, 2010, pp. 1706-1712.

[39] C. Lee, F. Reid, A. McDaid and N. Hurley, "Detecting highly overlapping community structure by greedy clique expansion", arXiv preprint:1002.1827, 2010.

[40] P. Pons and M. Latapy, “Computing communities in large networks using random walks”, International Symposium on Computer and Information Sciences, 2005, pp. 284-293.

[41] M. Rosvall and C. Bergstrom, "Maps of random walks on complex networks reveal community structure", Proceedings of the National Academy of Sciences, Vol. 105, 2008, pp. 1118-1123.

[42] D. Jin, B. Yang, C. Baquero, D. Liu, D. He and J. Liu, "A Markov random walk under constraint for discovering overlapping communities in complex networks", Journal of Statistical Mechanics: Theory and Experiment, 2011, p. P05031.

[43] T. Evans and R. Lambiotte, "Line graphs, link partitions, and overlapping communities", Physical Review E, Vol. 80, 2009, p. 016105.

[44] Y.Y. Ahn, J.P. Bagrow and S. Lehmann, "Link communities reveal multiscale complexity in networks", Nature, Vol. 466. 2010, pp. 761-764.

[45] G. Palla, I. Farkas, P. Pollner, I. Derenyi, and T Vicsek, "Directed network modules", New Journal of Physics, Vol. 9, 2007, p. 186.

[46] I. Farkas, D. Ábel, G. Palla and T. Vicsek, "Weighted network modules", New Journal of Physics, Vol. 9, 2007, p. 180.

[47] J. Kunegis, “Konect: the koblenz network collection”, 22nd International Conference on World Wide Web, 2013, pp. 1343-1350.

[48] J. Leskovec and R. Sosič, "SNAP: A general-purpose network analysis and graph-mining library", ACM Transactions on Intelligent Systems and Technology, Vol. 8, 2016, pp. 1-20.

[49] M. Bastian, S. Heymann and M. Jacomy, “Gephi: open source software for exploring and manipulating networks”, in Proceedings of the International Conference on Weblogs and Social Media, Vol. 8, 2009, pp. 361-362.

[50] J. Leskovec, K. Lang and M. Mahoney, “Empirical comparison of algorithms for network community detection”, in Proceedings of the 19th international conference on World Wide Web, 2010, pp. 631-640.


www.SID.ir